# A New Model for the Estimation of Cell Proliferation Dynamics Using CFSE Data

H.T. Banks, Karyn L. Sutton and W. Clayton Thompson
Center for Research in Scientific Computation and Center for Quantitative Sciences in Biomedicine
North Carolina State University, Raleigh, NC 27695-8212

Gennady Bocharov
Institute of Numerical Mathematics, RAS, Moscow, Russia

Marie Doumic
INRIA Rocquencourt, Projet BANG, Domaine de Voluceau, 78153 Rocquencourt, France

Tim Schenkel[1], Jordi Argilaguet[2], Sandra Giest[2], Cristina Peligero[2] and Andreas Meyerhans[1,2]
[1] Department of Virology, Saarland University, D-66421 Homburg, Germany
and
[2] ICREA Infection Biology Lab, Dept of Experimental and Health Sciences, Univ. Pompeu Fabra, 08003 Barcelona, Spain

August 20, 2011

**Abstract**

CFSE analysis of a proliferating cell population is a popular tool for the study of cell division and division-linked changes in cell behavior. Recently [13, 45, 47], a partial differential equation (PDE) model to describe lymphocyte dynamics in a CFSE proliferation assay was proposed. We present a significant revision of this model which improves the physiological understanding of several parameters. Namely, the parameter $\gamma$ used previously as a heuristic explanation for the dilution of CFSE dye by cell division is replaced with a more physical component, cellular autofluorescence. The rate at which label decays is also quantified using a Gompertz decay process. We then demonstrate a revised method of fitting the model to the commonly used histogram representation of the data. It is shown that these improvements result in a model with a strong physiological basis which is fully capable of replicating the behavior observed in the data.

**Key words:** Cell proliferation, cell division number, CFSE, label structured population dynamics, partial differential equations, inverse problems.

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

1. REPORT DATE
**20 AUG 2011**

2. REPORT TYPE

3. DATES COVERED
**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE
**A New Model for the Estimation of Cell Proliferation Dynamics Using CFSE Data**

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
**North Carolina State University,Center for Research in Scientific Computation,Department of Mathematics,Raleigh,NC,27695-8212**

8. PERFORMING ORGANIZATION REPORT NUMBER
**CRSC-TR11-05**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

**CFSE analysis of a proliferating cell population is a popular tool for the study of cell division and divisionlinked changes in cell behavior. Recently [13, 45, 47], a partial differential equation (PDE) model to describe lymphocyte dynamics in a CFSE proliferation assay was proposed. We present a significant revision of this model which improves the physiological understanding of several parameters. Namely, the parameter used previously as a heuristic explanation for the dilution of CFSE dye by cell division is replaced with a more physical component, cellular autofluorescence. The rate at which label decays is also quantified using a Gompertz decay process. We then demonstrate a revised method of fitting the model to the commonly used histogram representation of the data. It is shown that these improvements result in a model with a strong physiological basis which is fully capable of replicating the behavior observed in the data.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **33** | |

# 1 Introduction

The quantitative analysis of cell division is an important problem at the intersection of biology and mathematics. Of the myriad applications and active areas of research, meaningful quantification of lymphocyte dynamics associated with clonal expansion during an immunoassay constitutes a significant step toward understanding the complex underlying processes of the biological system. Understanding cell proliferation is important in numerous applications, from cancer and infectious disease diagnosis and treatment to immunosuppression therapies for transplant patients. These applications depend upon the accurate characterization of the the rates at which cells divide, differentiate, and die and, of equal importance, how changing intra- and extracellular conditions affect these rates. Most commonly, the proliferative characteristics of a cell population are measured in terms of the number of cells having undergone a specified number of divisions, as well as any division-linked changes which are observed. Thus the problem is two-fold. First, there is a need for an experimental procedure which can quickly and accurately provide division-related information in a population of dividing cells. Second, there is a need for mathematical models which can describe the data obtained from such a procedure.

Unlike many other cell types, the inherent mobility in vivo and nonadherence in vitro of lymphocytes makes the accurate determination of lineage very challenging [49] (although new techniques have been developed for this purpose [35]). In the absence of such data, one can look instead at the total number of divisions a cell has undergone since activation and how cells in different generations differ in phenotype (regardless of exact lineage). In the past two decades, a number of different techniques have been used for the study of cell growth and division [54, 66]. Early techniques, such as tritiated thymidine or bromodeoxyuridine (BrdU) uptake, while providing information regarding the fraction of dividing cells, are dependent upon cellular activation and do not provide information regarding how many divisions cells have undergone [50]. Lipophilic dyes which are incorporated into cellular membranes, such as PHK26, have been used successfully for the study of cell division history, although the uneven partitioning of the dye during mitosis can result in subsequent generations which are hard to distinguish [54, 66].

Since it was first described in 1994 [50], serial dilution of the fluorescent dye carboxyfluorescein succinimidyl ester (CFSE) has become the de facto method for the determination of such cellular division histories. CFSE is introduced into a culture of cells as carboxyfluorescein diacetate succinimidyl ester (CFDA-SE) which freely diffuses across the cell membrane and inside the cells. The acetate groups are then removed by intracellular esterases resulting in highly fluorescent CFSE which is less membrane permeant [54, 56]. CFSE is nonradioactive and stably incorporated (so that measurable concentrations of CFSE remain within a viable cell for several weeks in vivo); it provides quick, bright, and approximately uniform labeling of all cells in a population (regardless of cell type or activation) [50, 66]. With a peak absorption at 491nm and a peak emission at 517nm, CFSE is compatible with standard fluorescein cytometry setups [54, 56]. Using a flow cytometer, the fluorescence intensity (a surrogate for CFSE content) of individual cells can be measured. Because the CFSE content of a cell is divided approximately in half each time the cell divides, the number of divisions a cell has undergone can then be determined by comparing the measured amount of CFSE to the original CFSE content of an undivided cell [49, 50, 54, 56]. When individual cell fluorescent intensity measurements are binned into a histogram, each generation of cells appears as a “peak” in the histogram data.

Numerous protocols for the application of CFSE-based proliferation assays are available, and these protocols can be tailored to the specific goals of the experimenter [49, 50, 56, 68]. In particular, the compatibility of CFSE with other dyes renders possible the simultaneous measurement of division history and many other quantities, such as surface marker expression, cytokine content, and gene expression [49, 50, 66]. It is also possible to quantify the effects of extracellular conditions such as stimulation strength and duration on proliferative behavior [24, 30]. Certainly, a complete mathematical description of a lymphocyte response must eventually be able to account for such dynamic intra- and extracellular conditions. In this report, we postpone these additional complexities and focus on simplified linear models for cell division and death. Even in this simplified framework, we find that the resulting model has profound implications for the interpretation of cytometry data with a CFSE-based assay.

Because the flow cytometer provides measurements of individual cells within a sample, it is possible to obtain basic information regarding the proliferative capacity of a sample of cells simply by computing the proportion of dividing cells in a population. However, such descriptive and semi-quantitative methods are generally restricted to populations which divide synchronously [66, 68]. A more quantitative analysis is possible if the peaks in the histogram data (see Figure 1) are fitted with gaussian or log-normal curves to determine the numbers of cells in each generation [49, 56]. While these basic frameworks provide an efficient measurement of the gross behavior of a proliferating culture, a more complete analysis (as well as comparisons among different cultures and extracellular conditions) requires a more extensive mathematical framework to establish a quantitative measure of the proliferation dynamics of the population.

Detailed mathematical analysis for asynchronously proliferating populations of cells began in earnest with the work of Gett and Hodgkin [30]. By fitting CFSE histogram data with a series of log-normal curves, they computed cell numbers for each generation over the course of several days and tracked how these numbers changed over time. It was shown that cell numbers as a function of generation were well described by a gaussian curve, following the hypothesis that the primary source of asynchrony within the population was a normally distributed time-to-first-division. The authors go on to compute parameters such as mean division rate and mean time to first division. A revision of this model [24] incorporates cell death in the undivided population and the percentage of cells stimulated to divide. Because these models establish parameters which are readily identifiable from data sets, changes in parameters resulting from changing experimental conditions have been used to describe in exact terms how changing stimulatory and costimulatory conditions directly affect proliferative capacity [24, 30].

Other models to describe cell population dynamics have focused, to varying degrees, on mathematical formalisms of cell cycle dynamics. Several authors have used linear compartmental ordinary differential equations (ODEs) [57, 65] to predict the number of cells in each generation as a function of time. More commonly, the Smith-Martin [61] model of the cell cycle is used as the basis for a mathematization of cellular dynamics. In the Smith-Martin model, the cell cycle is divided into a stochastic A state and a deterministic B state (corresponding approximately to the G1 and S-G2-M phases of the cell cycle, respectively). In a differential equations framework, each generation of cells has compartments A and B; cells exit the A compartment with a known rate and remain in the B compartment for a fixed amount of time. This results in a system of delay differential equations [22, 23, 29, 53]. It has been shown that the early models of Hodgkin et al., [24, 30] can be described in terms of differential equations models [22, 43]. Several reports comparing autonomous and delay differential equations models have concluded that the delay term is vital to the accurate modeling of CFSE proliferation assay data because it prevents "rapid division cascades" by establishing a minimum cell cycle clock [22, 23, 29]. Alternatively, Asquith et al., [2] show that an autonomous differential equation model with a sufficiently small rate of division can accurately mimic the behavior of a delay system. Thus it seems clear that the interpretation of estimated parameters must be carefully done in the context of the particular model being used [23]. More recently, a delay differential equation model has been applied to cellular differentiation as well [59].

There are alternatives to differential equation models of cell-cycle based cell proliferation dynamics. The cyton model, initially proposed by Hawkins et al., [33], assumes that the cellular controls of growth/division and death are independent of one another. A particular cell in a given generation is assumed to have a fixed time-to-divide and time-to-die, both chosen from fixed probability distributions. Whichever parameter is smaller determines the fate of the cell in that generation; with each division, these two parameters are reset. Thus the behavior of the entire population is described by the probability distributions from which these two parameters are drawn. This has been shown to outperform the early ODE-compatible models [24, 30] and software implementing the method is freely available [34]. A recent analysis has shown that the cyton model is consistent with a Smith-Martin delay differential equation model under certain assumptions [41]. A generalization of the cyton model [25] has been formulated to account for recently observed correlations between the proliferative behavior of sibling and/or "cousin" cells.

Another approach to the cell-cycle based modeling of a proliferating population is the use of a stochastic Bellman-Harris type branching process [37, 38, 62, 69]. Branching process models are similar to the cyton model in that the fate of each cell is stochastic with fixed probability distributions for division and death. Hyrien and Zand [37] find that branching models form a superset of Smith-Martin based models, and analysis by Subramanian et al., [62] find it to be consistent with the cyton model.

Each of the models discussed thus far have been effectively used to provide various measures of the proliferative capacity of a population of cells. In general, these models are based upon estimation from the cell numbers computed by fitting CFSE histogram data with normal or log normal curves. Such approaches are straightforward and easy to implement, and the resulting cell numbers provide an accurate description of the static distribution of cells in the population. However, the imposition of particular shapes for the generational structure of CFSE histogram data can introduce biased insight into the generation structure of the cells, and hence into the resulting division and death rates. Alternatively, we propose that there is information to be learned not only from modeling the total numbers of cells, but also from the direct modeling of the complete experimental process. This is a more fundamental level of analysis of the kinetics of cell turnover which we believe to provide a more accurate assessment of the biological processes occurring in the population. Given such a goal, the development of CFSE and flow cytometry proliferation assays makes structured population models a natural framework in which to work. Significant literature exists on the subject of structured population models, going back at least as far as the Sinko-Streifer [60] model for general populations or the Bell-Anderson [17] model for volume-structured cell populations. More recently, "physiologically-structured" population models [52] have been developed for cell populations structured by age [1, 16, 26], cyclin content [16], and size [27, 55] as well as DNA-content [15].

The measurement of CFSE fluorescence intensity (FI) by a flow cytometer makes measured fluorescence intensity a natural structure variable for a structured population model. While not a physiological variable, the notion that such a structure might be used to accurately model cytometry data by accounting for the natural dilution of label was proposed at least as early as the year 2000 for BrdU-based assays [18]. To our knowledge, Luzyanina, et al., [47], proposed the first model to explicitly employ fluorescence intensity as a structure variable in a partial differential equation (PDE) framework. There it was shown that such a model can be effectively used for the tracking of a proliferating lymphocyte population stained with CFSE, and that such a model is as effective as compartmental ODE models for estimating the numbers of cells having undergone a specified number of divisions. The key idea behind the use of FI as a structure variable is that, because CFSE FI decreases upon division, fluorescence intensity can be used as a surrogate for division number. More recent work [6, 13, 45] has consistently demonstrated that this PDE framework can accurately model the observed histogram data from a CFSE-based proliferation assay. We believe that the primary benefit of using such a model lies in its ability to treat the measured FI data directly, thus accounting for the intracellular dynamics of label dilution while simultaneously estimating proliferation and death dynamics at the population level. Moreover, this method relies less on distinct peak separations in the CFSE histogram data, a potential advantage when working with heterogeneous cell populations.

While these models are indeed effective, the parameter estimates which resulted from fitting these models to an available data set seemed to suggest that label was being created during the process of cell division, a known impossibility [13]. In this document, we revisit the work presented in [13] and [47] primarily focusing on two key problems addressed there but not resolved. First, we address the issue with the apparent creation of label during cell division. It is shown that this apparent physiological impossibility is actually readily explained and removed from the models by the inclusion of cellular autofluorescence. Second, we investigate the functional form for the rate at which label naturally decays. By examining data from cells which were stained with CFSE but not stimulated to divide, we find that a minor modification of the exponential decay first proposed in [47] can provide a superior fit to the data. These two revisions (autofluorescence and biphasic label decay) provide important insights into the mathematical analysis of turnover kinetics for cells stained with CFSE and measured via flow cytometry. Their accurate modeling is vital to the meaningful estimation of population proliferation and death rates in a manner which is unbiased and mechanistically sound. Significantly, this new model is still sufficiently general to apply to a wide range of cell types and stimulation conditions and as such, might be used in a diagnostic setting [28] (e.g., to distinguish between healthy and diseased or abnormal cells based upon estimated proliferation rates).

## 2 CFSE Data Set

For the analysis here we use the same data set as in [13, 47]. This data set is the result of an in vitro proliferation assay with human blood mononuclear cells (PBMCs) taken from a healthy blood donor. Briefly, approximately $5\times10^6$ to $5\times10^7$ were stained with $5\mu M$ CFDA-SE and stimulated to divide with 2.5 $\mu g/mL$ phytohaemagglutinin (PHA). The cells are placed in well plates at a density of $1 \times 10^6$ cells per milliliter of RPMI-1640/10% fetal calf serum (FCS) nutrient medium. Every 24 hours, cells from a single well are harvested and transferred to Trucount tubes containing 51466 beads. These cells are then stained with fluorescently labeled anti-CD4 antibodies and analyzed via flow cytometry. More detailed information on the experimental protocol used for this particular data set can be found in [13, 47]. More information regarding the protocol in general can be found in [49, 50, 56, 66, 68].

At each sample time, a fraction of the population of cells placed in the Trucount tube and stained with fluorescently labeled antibodies are counted by the cell sorter. This subpopulation contains all cell types present in a PBMC culture (CD4+ and CD8+ T cells, B cells, monocytes, etc.), however only CD4+ T cells are considered for mathematical analysis after gating them based on size, granularity and CD4+ expression. Because the entire contents of the tube are not collected, the cell counts obtained from the cytometer are scaled upward by the ratio

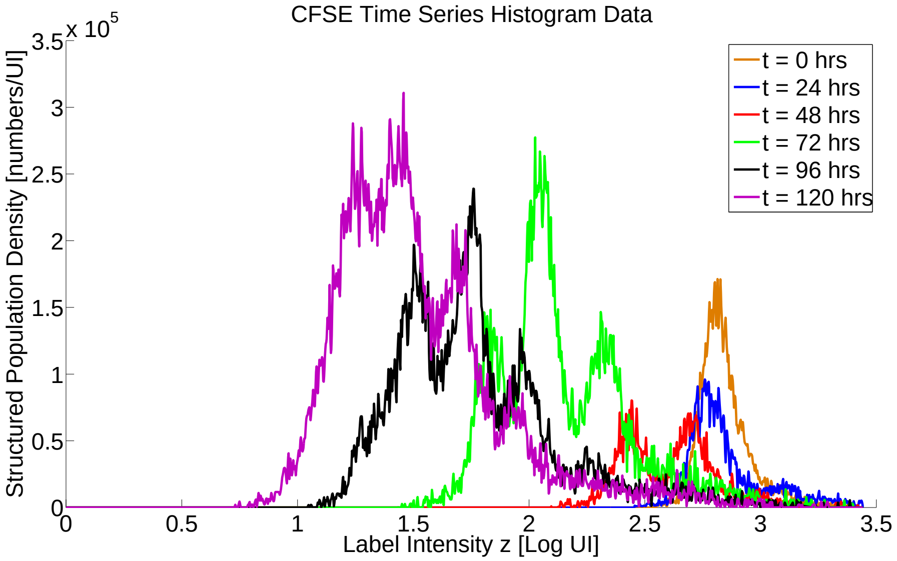

Figure 1: Original histogram data from [13, 47].

of known beads in the tube to the number of beads actually counted. We remark that, while this report focuses exclusively on human CD4+ lymphocytes cultured in vitro, CFSE-based assays have been used successfully in a wide variety of applications, including other T lymphocytes, B lymphocytes, NK cells, bacteria, fibroblasts, hematopoietic stem cells, and smooth muscle cells [49, 54, 56].

Qualitatively, the flow cytometer returns a measure of the fluorescence intensity of a given cell, owing primarily to the presence of CFSE within the cell. In order to obtain this measurement, the flow cytometer uses hydrodynamic focusing to push cells one at a time through a beam of laser light. This light is absorbed and then emitted again by the electrons in CFSE. This emitted light is filtered and then quantified by a photometer. While the measurement process itself is complex, one should note that it is possible to measure thousands of individual cells in a matter of seconds (so that it is reasonable to assume that a sample does not undergo any changes during the measurement process). It is known that measured CFSE FI has a linear relationship with the concentration of CFSE used in the staining process [49, Fig. 3] and is expected to correlate with the mass of CFSE within a cell. Because CFSE FI divides approximately in half with each subsequent division (at least for the first few generations; see Section 3) it is most convenient to use a logarithmic scale for CFSE FI.

The most common representation for CFSE FI data is a series of histograms. (While it is possible, in general, to choose the bins for the histograms as one wishes, we remark that the current data set was already reported as histogram data when we began efforts on it.) The current data set is shown in Figure 1. It is this histogram data for which we seek a mathematical model. Each peak in the histogram data consists of cells which have undergone the same number of divisions. Over time, all cells (even in the absence of division) slowly drift to the left, reflecting a loss of label. An effective mathematical model must adequately describe both the emergence of these distinct peaks as well as the slow decay of the label.

While the results in [13] were generated by fitting to the data at all five points in time ($t$ = 24, 48, 72, 96, and 120 hours), the current study will not make use of the data at $t = 72$ hours. Because CFSE is added to the cell culture at the beginning of the experiment but not afterward, the total mass of CFSE in culture cannot increase over the course of the experiment. This mass can decrease as a result of cell death and the natural decay/catabolism of CFSE within a cell. (While the separate measurements in time are obtained from distinct populations of cells in separate wells, the assumption that each well contains a sufficiently similar population is standard.) Because fluorescence intensity is approximately proportional to the mass of CFSE within a cell, the sum of all cells in a population, weighted by measured FI, provides an indication of the mass of CFSE within the measured population. We have found that this ‘total label content’ (the sum of all cells in the histogram, weighted by the measured FI) is greater at $t = 72$ hours than at the previous time point (see Figure 2), indicating

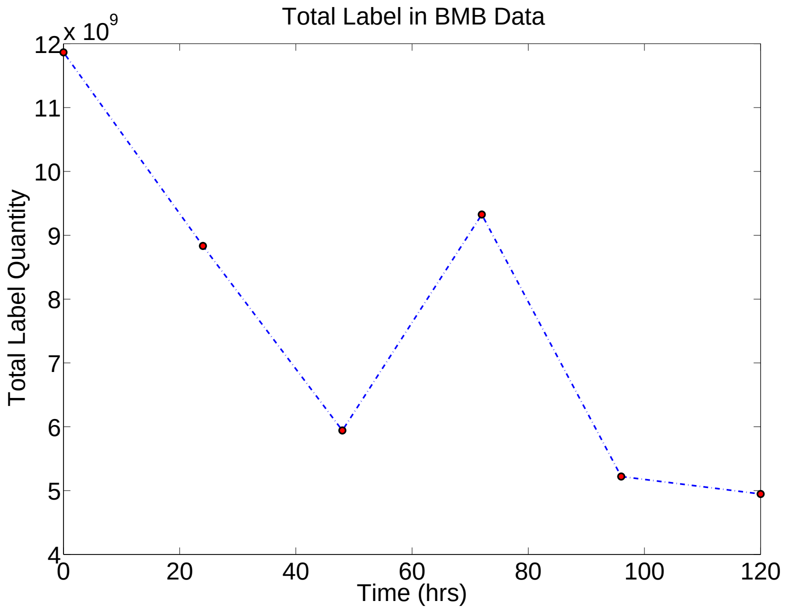

Figure 2: Total label content $\int zn(t,z)dz$ over time for the data from [13]. The increase at $t = 72$ hours is a physiological impossibility.

a net increase in the mass of CFSE between $t = 48$ hours and $t = 72$ hours, a physical impossibility. The causes of this increase are not currently known. It is possible that this unusual result is explained by naturally occurring fluctuations in the number of beads counted by the cytometer, or possibly by the presence of additional cells types (e.g., monocytes) in the data. This unusual feature is present (at various measurement times) in several additional data sets which have been collected. As such, this feature will need to be addressed in the future by a more accurate observation model. At any rate, we assume for the current report that this anomaly is the result of measurement or scaling error or some unknown and unmodeled biological event; hence the $t = 72$ hours data point will not be included in the present investigation. The new mathematical model proposed in this report, as well as the mathematical model from [13], have both been calibrated with and without the data point at $t = 72$ hours. The results (unpublished) confirm our suspicion as both models (which are derived from CFSE mass-conservation principles) are incapable of accounting for the erroneously large quantity of CFSE observed at $t = 72$ hours.

## 3 Mathematical Model

We begin this section by first recalling [13, 47] the previous PDE models proposed to describe the data. Let $n(t,z)$ be the label-structured population density indicating the density of cells at time $t$ (hours) and log label intensity $z$ (log units of intensity, log UI). In the analysis of [13], it was shown that the most effective mathematical model had the form

$$\begin{aligned} \frac{\partial n(t,z)}{\partial t} + \frac{\partial [cn(t,z)]}{\partial z} &= -(\alpha(t,z+ct) + \beta(z+ct))n(t,z) \\ &+ \chi_{[z_{\min}, z_{\max} - \log\gamma]} 2\gamma\alpha(t, z+ct+\log\gamma)n(t, z+\log\gamma), \end{aligned} \tag{1}$$

where $\alpha(t,s)$ is the rate of cell proliferation (hr$^{-1}$), $\beta(s)$ is the rate of cell death (hr$^{-1}$), $c$ is the rate at which label is lost (log UI hr$^{-1}$), and $\gamma$ is the ratio of CFSE FI of a mother cell to that of a daughter cell. Thus the second term on the left represents velocity of decay of the florescence label while the last term on the right represents rate of production of new cells due to cell division. A complete derivation and detailed explanations are given in [6, 13].

### 3.1 Physiological Interpretation of $\gamma$

It was shown in [13] that the model (1) is quite capable of providing an accurate fit to the data set at hand. However, the parameter $\gamma$ was used to represent an unknown process responsible for determining the CFSE FI of two daughter cells given the CFSE FI of a mother cell. It was not known at the time what processes might be represented by $\gamma$ or how that parameter should be interpreted (see also [47] where $\gamma$ was first introduced). We would like to make this model more physically relevant by explaining the mechanism underlying the parameter $\gamma$.

Mathematically, the parameter $\gamma$ determines the peak-to-peak separation between subsequent generations of cells (i.e., each generation has a CFSE FI approximately $\log_{10}\gamma$ less than the previous generation in the log FI coordinate $z$). Given the definition of $\gamma$ as the ratio of CFSE FI of a mother cell to that of a daughter cell, it is expected that $\gamma \geq 2$, with $\gamma > 2$ if label is lost during the process of cell division. However, the best fit parameter from [13] was $\gamma^* = 1.5169$, implying the creation of label at division. Similar results were also obtained in [47]. Indeed, forward simulations of the above model demonstrate that $\gamma = 2$ is significantly too large to fit the given data set, regardless of the values assigned to other parameters.

One possible solution conjectured in [13] to explain this discrepancy was that the measurement of CFSE FI may be indicative of the concentration of CFSE, rather than its mass. The observations, then, would represent an effective integration over the various cell volumes present in the data. While appealing, this explanation does not appear to be the case. Physically, one expects that measured CFSE FI would depend on the number of CFSE molecules within the cell, and hence on the mass of CFSE. Indeed, when cells stained with CFSE are introduced to a stimulating agent, the cells quickly increase in size (thus decreasing the CFSE concentration), but the measured CFSE FI is essentially unchanged [50, Fig. 6].

We propose here an alternative solution to this apparent $\gamma$-related dilemma. While it is often stated that the subsequent peaks in the CFSE histogram data are evenly spaced [49], close observation reveals that this is not actually the case. Although the peaks corresponding to low division numbers (Generations 0, 1, 2, 3) are approximately evenly spaced, peaks corresponding to larger division numbers are closer and closer together [50, Fig. 1]. In other words, the parameter $\gamma$ appears to change with division number.

These observations can be collectively explained by the consideration of cellular autofluorescence and its effects on the measurement process. As discussed in Section 2, the flow cytometer measurement process uses light as a surrogate for CFSE content. However, not all light incident upon the photodetector is the result of emission from CFSE molecules. All cells, even those unstained with CFSE, have a natural brightness and will give off small but detectable amounts of light. We assume here that this feature, the *cellular autofluorescence* does not change as cells divide and does not decay slowly like CFSE fluorescence. It may vary with time for other reasons [3], but we ignore this in our initial treatment.

Let $X_i$ be the total measured FI of a single cell, measured when that cell has undergone $i$ divisions. (The use of the capital letter is meant to distinguish this discrete quantity from the continuous state variable to be used in the revised model derived below.) Under the assumption that cellular autofluorescence intensity (AutoFI) and CFSE FI are additive, then the total fluorescence intensity of a cell is

$$X_i = X_i^{\mathrm{CFSE}} + X^{\mathrm{Auto}}. \tag{2}$$

Because AutoFI does not change as a cell divides, it follows that this cell with intensity $X_i$ will generate two cells in the next generation, each of which has total FI

$$X_{i+1} = X_i^{\mathrm{CFSE}}/2 + X^{\mathrm{Auto}}. \tag{3}$$

Contrary to previous interpretations of the parameter $\gamma$, one can see from Equation (3) that it is actually expected that the ratio of total FI of a mother cell to that of a daughter cell is expected to be less than 2. Moreover, provided $X_i^{\mathrm{CFSE}} >> X^{\mathrm{Auto}}$, this ratio is approximately equal to two. With each division, $X_i^{\mathrm{CFSE}}$ decreases and the ratio decreases; as $X^{\mathrm{Auto}}$ accounts for a larger and larger percentage of the total measured FI, the ratio decreases quicker and quicker until $X_i^{\mathrm{CFSE}} \approx 0$ and the ratio of mother-to-daughter intensities is approximately 1. Thus it appears that cellular autofluorescence is sufficient to account for the observed relationships between subsequent division peaks in the data. Indeed, this is shown to be the case in Section 5.

We remark that the phenomenon of autofluorescence when using fluorescent dyes to study biological materials is not particularly new. In fact, the role of AutoFI described above was acknowledged specifically for CFSE data sets as early as 1996 [36], and a formula corresponding to (3) above appears in [49]. However, autofluorescence has not been used in previous PDE formulations [13, 45, 47] to describe the dilution of CFSE by division. Thus the incorporation of AutoFI into the mathematical model presented below is an important revision to the physiological basis of the PDE model.

## 3.2 Revised Model Derivation

At this point, we pause momentarily in order to revise the PDE model derivation from the Appendix of [13]. We do so to provide a general framework with which to build additional improvements to the model and inverse problem procedure. As before, the derivation follows the mass-balance principles of the Bell-Anderson [17] and Sinko-Streifer [60] models.

Let $n(t,x)$ be the structured population density of a population of cells labeled with CFSE, where $t$ is time (in hours) and the structure variable $x$ is a given fluorescence intensity (FI) of a cell (in arbitrary units of intensity, UI). Then

$$N(t) = \int_{x_0}^{x_1} n(t,x)dx. \tag{4}$$

represents the total number of cells with fluorescence intensity in $(x_0, x_1)$ at time $t$. Here $x_0$ and $x_1$ are arbitrary. Let $\Delta x(t,x,\Delta t)$ be the average increase of FI of cells with initial intensity $x$ during the interval $(t, t+\Delta t)$ and assume that $\Delta t$ is chosen such that $|\Delta x| << x_1 - x_0$ (so that the number of cells which move into the region via division and subsequently divide, die, or drift out of the region is negligible). It should be noted that $\Delta x$ will be non-positive (as cells cannot increase in FI). Thus subtraction by $\Delta x$ actually results in a larger value. While counterintuitive, this definition is maintained in order to harmonize with other structured population models.

Consider the change in $N(t)$ during the time interval $(t, t+\Delta t)$, i.e., the quantity $N(t+\Delta t) - N(t)$. Five possible contributions will be considered:

(i.) Cells with intensity in the interval $[x_1, x_1 - \Delta x(t, x_1, \Delta t)]$, losing FI according to $\Delta x$:

$$\int_{x_1}^{x_1 - \Delta x(t,x_1,\Delta t)} n(t,x)dx.$$

(ii.) Cells with intensity in the interval $[x_0, x_0 - \Delta x(t, x_0, \Delta t)]$, losing FI according to $\Delta x$:

$$\int_{x_0}^{x_0 - \Delta x(t,x_0,\Delta t)} n(t,x)dx.$$

(iii.) Cells which would have contributed to $N(t+\Delta t)$ had they not died:

$$\int_t^{t+\Delta t} \int_{x_0 - \Delta x(t,x_0,t+\Delta t-\tau)}^{x_1 - \Delta x(t,x_1,t+\Delta t-\tau)} \beta(\tau,x)n(\tau,x)dxd\tau.$$

(iv.) The disappearance of cells from the region due to proliferation:

$$\int_t^{t+\Delta t} \int_{x_0 - \Delta x(t,x_0,t+\Delta t-\tau)}^{x_1 - \Delta x(t,x_1,t+\Delta t-\tau)} \alpha(\tau,x)n(\tau,x)dxd\tau.$$

(v.) The gain of daughter cells (two of them) in the region as a result of proliferation in the parent region:

$$\chi_{[x_a,x^*]} 2 \int_t^{t+\Delta t} \int_{2(x_0 - \Delta x(t,x_0,t+\Delta t-\tau)) - x_a}^{2(x_1 - \Delta x(t,x_1,t+\Delta t-\tau)) - x_a} \alpha(\tau,x)n(\tau,x)dxd\tau,$$

where $x^* = x_{\max}/2 + x_a$ and $x_a$ is the natural autofluorescence of unstained cells.

We remark that $\alpha$ and $\beta$ are the rates of cell proliferation and death, respectively, with units hr$^{-1}$. It follows that the difference $N(t+\Delta t) - N(t)$ is the sum of the components (i.) and (v.) less the contributions of components (ii.) - (iv.). Following the standard procedure of dividing by $\Delta t$ and letting $\Delta t \to 0$, we obtain $\frac{dN}{dt}$ on the left side of the equation. We now treat the right side of the equation term by term.

For the first term on the right side, if $n(t,x)$ is continuous in $t$ and $x$ (a reasonable assumption), the mean value theorem (MVT) implies that there exists a $\theta \in [x_1, x_1 - \Delta x(t,x_1,\Delta t)]$ such that

$$\int_{x_1}^{x_1-\Delta x(t,x_1,\Delta t)} n(t,x)dx = -\Delta x(t,x_1,\Delta t)n(t,\theta).$$

Assuming $\Delta x$ is continuous in $\Delta t$ (that is, there is no instantaneous label loss) and varies smoothly,

$$\lim_{\Delta t\to 0} \frac{-\Delta x(t,x_1,\Delta t)}{\Delta t} n(t,\theta) = -v(t,x_1)n(t,x_1). \tag{5}$$

where we have defined $\frac{dx}{dt} = v(t,x)$, the instantaneous rate of FI change of cells with intensity $x$ and time $t$. Applying the same argument for the second term,

$$-\int_{x_0}^{x_0-\Delta x(t,x_0,\Delta t)} n(t,x)dx = v(t,x_0)n(t,x_0). \tag{6}$$

In the consideration of the third term, define

$$u_\beta(\tau) = \int_{x_0-\Delta x(t,x_0,t+\Delta t-\tau)}^{x_1-\Delta x(t,x_1,t+\Delta t-\tau)} \beta(\tau,x)n(\tau,x)dx.$$

Then if $\Delta x(t,x,\Delta t)$ and $\beta(\tau,x)n(t,x)$ are continuous functions of their variables, so is $u_\beta(\tau)$ and by the MVT, there exists a $\theta' \in [t, t+\Delta t]$ such that

$$\frac{1}{\Delta t}\int_t^{t+\Delta t} u_\beta(\tau)d\tau = u_\beta(\theta').$$

Thus it follows that

$$\lim_{\Delta t\to 0} u_\beta(\theta') = u_\beta(t) = \int_{x_0}^{x_1} \beta(t,x)n(t,x)dx, \tag{7}$$

assuming $\Delta x(t,x,0) = 0$ for all $t,x$ (which follows from the previous assertion regarding the smoothness of $\Delta x$ in $\Delta t$). Using a similar argument for the fourth term,

$$\lim_{\Delta t\to 0} u_\alpha(\theta') = u_\alpha(t) = \int_{x_0}^{x_1} \alpha(t,x)n(t,x)dx, \tag{8}$$

where $u_\alpha(\tau)$ has the obvious definition. For the final term, the same argument along with the change of variables $\xi = (x + x_a)/2$ results in

$$\chi_{[x_a,x^*]}2\lim_{\Delta t\to 0} u_{\tilde{\alpha}}(\theta') = \chi_{[x_a,x^*]}4\int_{x_0}^{x_1} \alpha(t,2x-x_a)n(t,2x-x_a)dx. \tag{9}$$

Altogether, we can assemble (5) - (9) to obtain

$$\begin{aligned}\frac{dN}{dt} &= -v(t,x_1)n(t,x_1) + v(t,x_0)n(t,x_0) - \int_{x_0}^{x_1} \beta(t,x)n(t,x)dx \\ &\quad - \int_{x_0}^{x_1} \alpha(t,x)n(t,x)dx + \chi_{[x_a,x^*]}4\int_{x_0}^{x_1} \alpha(t,2x-x_a)n(t,2x-x_a)dx.\end{aligned}$$

On the left side, differentiating $N(t) = \int_{x_0}^{x_1} n(t,x)dx$ with respect to $t$ results in

$$\frac{dN}{dt} = \int_{x_0}^{x_1} \frac{\partial n(t,x)}{\partial t}dx.$$

Finally, by applying the Fundamental Theorem of Calculus to the first two terms on the right side, simplifying and rearranging,

$$\begin{aligned} \int_{x_0}^{x_1} \frac{\partial n(t,x)}{\partial t} + \int_{x_0}^{x_1} \frac{\partial (v(t,x)n(t,x))}{\partial x} &= \\ - \int_{x_0}^{x_1} (\alpha(t,x)+\beta(t,x))n(t,x) + \chi_{[x_a,x^*]} 4 \int_{x_0}^{x_1} \alpha(t, 2x-x_a)n(t,2x-x_a). \end{aligned}$$

Equivalently (because $x_0$ and $x_1$ were arbitrary),

$$\begin{aligned} \frac{\partial n(t,x)}{\partial t} &+ \frac{\partial [v(t,x)n(t,x)]}{\partial x} = \\ &- (\alpha(t,x)+\beta(t,x))n(t,x) + \chi_{[x_a,x^*]} 4\alpha(t,2x-x_a)n(t,2x-x_a). \end{aligned} \tag{10}$$

We remark that the above derivation is not very different from that already presented in [13]. The key differences are the notational change in permitting the dependence of the proliferation and death rates ($\alpha$ and $\beta$) and the label loss rate ($v$) on both time $t$ and measured FI $x$. This model also explicitly incorporates the even division of CFSE between daughter cells while also accounting for the presence of cellular AutoFI.

### 3.3 Gompertz Decay of Label

Given the model (10), we now turn our attention to the label loss rate $v(t,x)$. Because the mathematical model estimates cell proliferation and death rates in terms of the CFSE FI structure variable (as a surrogate for division number), the manner in which CFSE naturally decays directly affects the cell turnover parameter estimates. Thus, our understanding of the underlying biology (in the form of cell proliferation and death rate estimates) is closely tied to the accurate modeling of label decay. In order to provide parameter estimates which are unbiased, it is of vital importance that the label loss rate $v(t,x)$ accurately reproduces the natural decrease in CFSE FI observed in the data.

It was hypothesized in [47] that an exponential rate of loss is sufficient to model the label loss observed in the data. In order to validate this assumption, a PBMC culture was taken from two donors and stained with CFSE following the standard procedure. However, these cells were not stimulated to divide. Because only viable cells are included when the cytometry data is gated, any decrease in mean FI in the population must be the result of natural CFSE FI decay. Over the course of 160 hours, cells from each donor were measured at 24 distinct time points in triplicate and the mean total FI of each sample was recorded. The data is given in Table 1 and is shown graphically in Figure 3.

We would like to determine what functional forms might be used in order to quantify the label loss observed in the data. Following the assumptions of [13, 47], we begin with a model of label loss that decays exponentially to the autofluorescence of unlabeled cells,

$$x_1(t) = (x(0) - x_a)e^{-ct} + x_a. \tag{11}$$

However, it appears from the data (particularly for Donor 1) that the rate of exponential decay of label may itself decrease as a function of time. This can be modeled by the Gompertz decay process [40, pg. 12]

$$x_2(t) = (x(0) - x_a)\exp\left(-\frac{c}{k}\left(1 - e^{-kt}\right)\right) + x_a. \tag{12}$$

The loss rate function, of vital importance to the PDE formulation (10), is $v(t,x) = \frac{dx}{dt}$. Thus the equations (11 - 12) correspond to the loss rate functions

$$v_1(x) = -c(x - x_a), \tag{13}$$

and

$$v_2(t,x) = -c(x - x_a)e^{-kt}, \tag{14}$$

respectively. We remark that (12) is a generalization of (11), the latter being the limiting value (as $k \to 0$) of the former. Thus, it would be ideal to fit both models to the data and use statistical tests to determine if the

| Time (hours) | Donor 1 | | | Donor 2 | | |
|---|---|---|---|---|---|---|
| 0.00 | 44311 | 43272 | 45369 | 40878 | 41593 | 41993 |
| 2.00 | 33782 | 37720 | 36961 | 36755 | 32585 | 25705* |
| 4.00 | 29331 | 30043 | 29634 | 30818 | 28565 | 27144 |
| 6.00 | 29235 | 28526 | 31283 | 26498 | 27666 | 26354 |
| 8.00 | 25899 | 27229 | 29839 | 25856 | 18404* | 25060 |
| 10.00 | 26651 | 27691 | 27406 | 24846 | 25336 | – |
| 12.50 | 29471 | 27610 | 27852 | 24172 | 25506 | 26290 |
| 19.25 | 27201 | 27254 | 24718 | 30272* | 26922 | 26937 |
| 21.25 | 27062 | 23601 | 25342 | 27436 | 27646 | 29527 |
| 23.25 | 20758 | 24967 | 24640 | 25680 | 26474 | 26482 |
| 25.25 | 26585 | 23512 | 23400 | 25947 | 26711 | 26379 |
| 27.25 | 23356 | 24882 | 22898 | 26040 | 25551 | 28393 |
| 29.25 | 23660 | 21729 | 24288 | 23975 | 22490 | 23471 |
| 31.50 | 22768 | 20914 | 21268 | 21153 | 21244 | 21759 |
| 33.25 | 23897 | 24198 | 24758 | 25337 | 24053 | 24749 |
| 50.25 | 22623 | 23504 | 24696 | 26138 | 25672 | 27361 |
| 54.75 | 21910 | 21120 | 21986 | 25777 | 24564 | 26069 |
| 59.25 | 21877 | 26290 | 23829 | 21932 | 21302 | 27558* |
| 77.25 | 22099 | 24160 | 21420 | 25769 | 24108 | 26511 |
| 85.75 | 20731 | 21827 | 22108 | 26604 | 26293 | 26306 |
| 98.75 | 21468 | 20993 | 21220 | 22869 | 21760 | 22579 |
| 123.75 | 18836 | 18887 | 18313 | 20369 | 20533 | 21003 |
| 131.25 | 20132 | 20119 | 20799 | 20908 | 22234 | 26666* |
| 150.75 | 22199 | 19822 | – | 26542 | 23417 | – |

Table 1: Data sets collected from Donor 1 and Donor 2, rounded to the nearest integer. Several outliers are noticeable in data from Donor 2 and have been marked with an asterisk. All data is given in arbitrary units of intensity (UI).

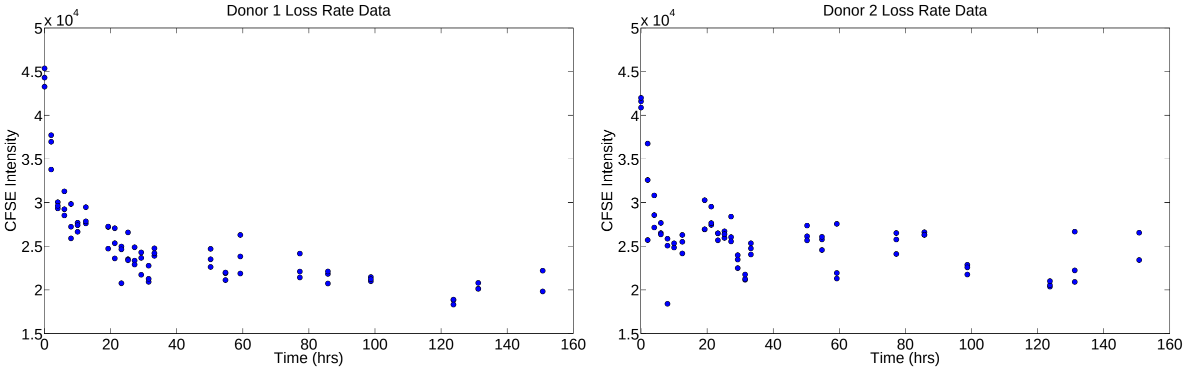

Figure 3: CFSE Intensity Data from unstimulated cells. Left: Donor 1. Right: Donor 2.

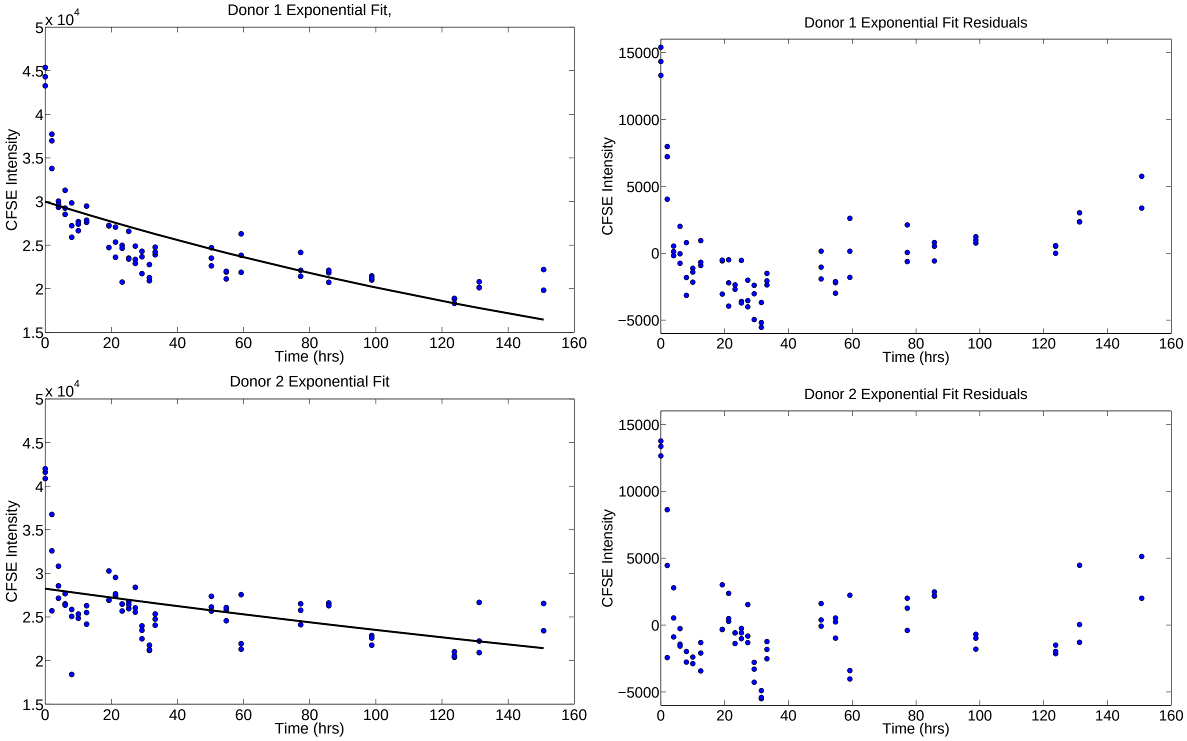

Figure 4: Results of fitting the exponential model (11) to the mean CFSE data in Table 1. For both Donor 1 (top) and Donor 2 (bottom), we see from both the fit to the data (left) and the residual plot (right) that the model is not capable of accurately replicating the observed data when $x_a$ is set to 50.

| Model | Donor 1 | Donor 2 |
|---|---|---|
| Exponential | $1.176515 \times 10^9$ | $1.073325 \times 10^9$ |
| Gompertz | $2.853192 \times 10^8$ | $4.285324 \times 10^8$ |

Table 2: OLS cost values for fitting the exponential (11) and Gompertz (12) to the CFSE decay data from two donors of Table 1. We see a very clear reduction in cost when using the Gompertz model.

model refinement is warranted. However, it is not possible to uniquely identify the parameter $x_a$ in either of the two models using an ordinary least squares procedure with only the data from Table 1. On the other hand, the parameter $x_a$ appears in other parts of the model (10), and it seems possible to conclude that this parameter would be readily identifiable once incorporated into (10) with the full proliferation data.

In order to at least begin to compare these two models, we set the parameter $x_a$ to the physiologically reasonable value of 50 in both models. (Values reported in the literature ranged from 10 to 100 [48, Figure 1], [50, Figure 2] and [56, Figure 3].) An ordinary least squares (OLS) procedure is then used to fit the remaining parameters ($c$ and $x(0)$ for the Exponential model, $c$, $k$ and $x(0)$ for the Gompertz model) for data from both donors. The results for the exponential model are shown in Figure 4 and the results for the Gompertz model are shown in Figure 5. Based upon the lines of best fit in comparison with the data, it seems clear that the Gompertz decay model more accurately describes the available data. This is confirmed by the comparison of cost given in Table 2.

A few additional comments are in order. First, we remark that, while the parameter $x(0)$ is included in the models (11) and (12) used to fit the CFSE label loss data sets, it is actually the loss rate function (either (13) or (14)) that will be included in the PDE model (10) and thus this parameter will not actually require estimation in the full PDE model. Also, when using CFSE-labeled cells for a proliferation assay, there is an additional hour of preparation time between CFSE labeling and the first measurement time, during which the cells are stimulated

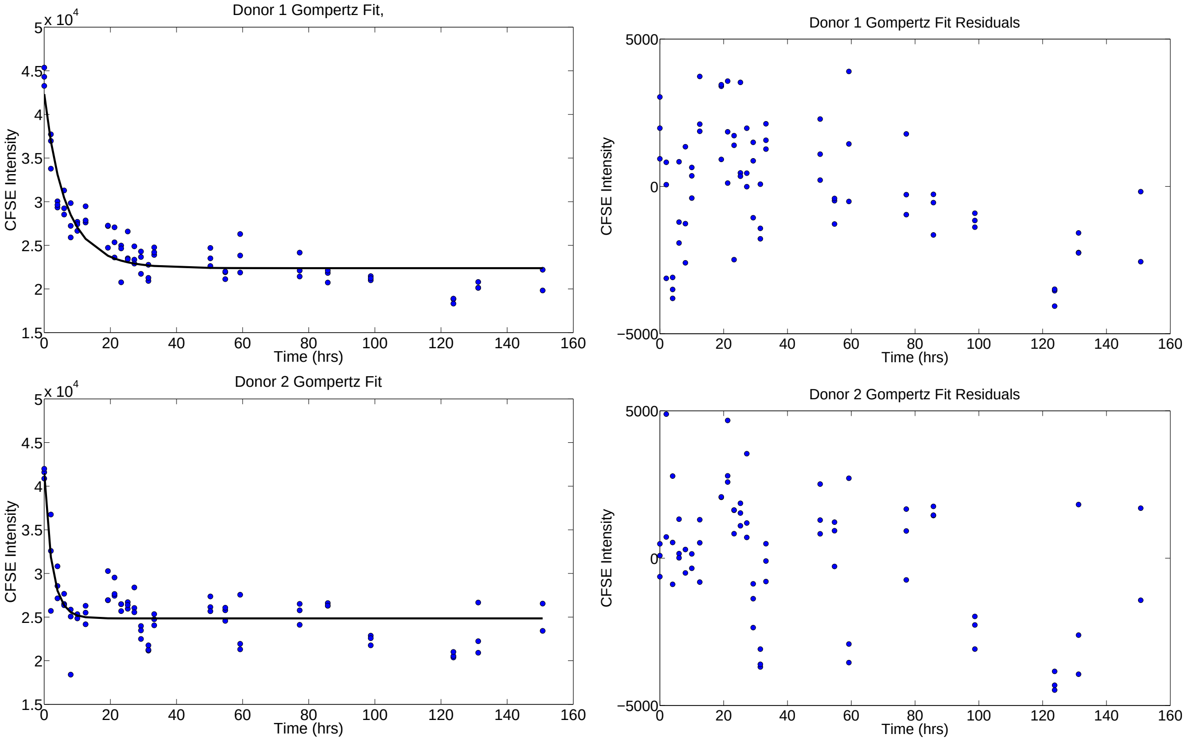

Figure 5: Results of fitting the Gompertz model (12) to the mean CFSE data in Table 1. We see a much improved fit to both donors when compared to the exponential fit of Figure 4. For both donors the parameter $x_a$ of (12) is set to 50.

to divide. Thus we would expect slightly different parameter estimates when either of these models is used with proliferation assay data. Moreover, because this additional hour of preparation time occurs during the first hour after staining, the rapid decay observed in both data sets in Figure 3 may be partially removed from the data. Because of these considerations, it is unclear exactly how much improvement to expect from using (14) as opposed to (13) in (10).

We also remark that the primary reason for the failure of the exponential model is the location of the equilibrium point in the data (as compared to that predicted by the model). We see from Equation (11) that the exponential model predicts $x \to x_a$ as $t \to \infty$. However, it is known that CFSE stained cells retain detectable fluorescence for up to several weeks in vivo [50]. The exponential model cannot accurately account for both the rapid decline in CFSE FI during the first few hours of staining and the slow decline once the label has been stably incorporated. (We can, for the current data sets, allow $x_a$ to attain values of approximately $2.2 \times 10^4$; in this case the exponential model fits the data at least as well as the Gompertz model. However, this is not a physiologically reasonable value of $x_a$.)

Physiologically , it is known that after the conversion of CFDA-SE to CFSE by intracellular esterases, CFSE can still exit the cell at a slow rate (compared to free diffusion). However, the succinimidyl group of CFSE reacts covalently with amines attached to intracellular proteins. While some of the resulting conjugates are short-lived (either because they exit the cell or are rapidly degraded) other conjugates are stably incorporated inside the cell and remain so for an extended period of time. These stable conjugates are decreased further only by the natural turnover of the intracellular proteins to which they are bound [54]. These processes combine to produce the commonly observed "biphasic decay" of CFSE FI over time [51, 66]. In other words, it seems necessary for the rate of CFSE FI (exponential) decay to decrease in time. This is precisely the feature of the Gompertz decay model [40].

# 4 Parameter Estimation Procedure

With the primary features of the revised model now addressed, we are ready to validate the model with data. The current model, accounting for CFSE AutoFI and the Gompertz decay of the label, is

$$\begin{aligned} \frac{\partial n(t,x)}{\partial t} &- ce^{-kt}\frac{\partial[(x-x_a)n(t,x)]}{\partial x} = \\ &- (\alpha(t,x)+\beta(t,x))n(t,x) + \chi_{[x_a,x^*]}4\alpha(t,2x-x_a)n(t,2x-x_a). \end{aligned} \tag{15}$$

At the right boundary ($x = x_{\max}$), we expect that there are no cells which can drift (via label loss) into the computational domain. At the left boundary, a zero flux condition is imposed to prevent cells from drifting to CFSE FI values less than the AutoFI of unlabeled cells. Thus the boundary conditions are

$$\begin{aligned} n(t,x_{\max}) &= 0 \\ v(t,x_a)n(t,x_a) &= 0. \end{aligned} \tag{16}$$

Note from (14) we have $v(t,x_a) = 0$ for all $t$, and hence the left boundary condition is trivially satisfied. Finally, we assume we are given some initial condition

$$n(0,x) = \Phi(x), \tag{17}$$

which is the initial distribution of cells as a function of FI.

## 4.1 Model Change of Variables

While the model (15) and its associated initial and boundary conditions suitably describe the dynamics for a CFSE labeled population of dividing lymphocytes, the model is not conducive to finite difference methods for numerical solutions. The CFL condition for stability requires

$$\Delta t < \frac{\Delta x}{\max|v(t,x)|} = \frac{\Delta x}{\max c(x-x_a)}. \tag{18}$$

Because $x_{\max} >> x_a$, the computational domain is quite large, and $\max(x - x_a) \sim 10000$. Moreover, this large domain must have a relatively fine mesh, as features of the solution become less distinguishable with increasing division number. We see that, given $c \sim 0.1$, $k \sim 0.001$, and $\Delta x \sim 0.1$ (all reasonable parameters), it would take upward of $10^5$ time steps to compute the solution out to $t = 120$ hours, and this must be done for $10^5$ points on the structure variable grid. Rather than attempt these expensive computations, we seek a change of variables that will lead to a faster numerical solution.

The most immediate choice is to use the change of variables $z = \log_{10} x$, as the data is given in this coordinate. While this change of variables was effective in [13, 47], it is less effective here because of the different form of the label loss rate function. Instead, we use the change of variables $y = \log_{10}(x - x_a)$. Then $x = 10^y + x_a$ and

$$\frac{dy}{dx} = \frac{1}{(x - x_a)\ln(10)}.$$

Let $\tilde{n}(t, y) = 10^y \ln(10) n(t, x(y)) = 10^y \ln(10) n(t, 10^y + x_a)$. We remark that the factor $10^y \ln(10)$ arises from the chain rule in the integral form of (15) and is needed to conserve the total label in the population. With this change of variables the new PDE model is

$$\begin{aligned} \frac{\partial \tilde{n}}{\partial t} - ce^{-kt} \frac{\partial}{\partial y}\left[\frac{\tilde{n}(t,y)}{\ln 10}\right] &= -(\tilde{\alpha}(t,y) + \tilde{\beta}(t,y) - ce^{-kt})\tilde{n}(t,y) \\ &+ \chi_{(-\infty,y^*]} 2\tilde{\alpha}(t, y + \log_{10} 2)\tilde{n}(t, y + \log_{10} 2), \end{aligned} \tag{19}$$

where $\tilde{\alpha}(t, y) = \alpha(t, x(y))$, $\tilde{\beta}(t, y) = \beta(t, x(y))$ and $y^* = y_{\max} - \log_{10} 2$. The new initial condition is $\tilde{\Phi}(y) = 10^y \ln(10)\Phi(t, 10^y + x_a)$. The right boundary condition follows immediately from (16) while the left boundary has been removed to $y = -\infty$. We remark that the CFL condition for the PDE (19) is

$$\Delta t < \frac{\Delta y \ln 10}{ce^{-kt}}$$

which is significantly easier to satisfy.

## 4.2 Parameterizations of $\alpha$ and $\beta$

We next turn our attention to the parameterizations of the functions $\alpha(t, y)$ and $\beta(t, y)$. Because our goal is the estimation of lymphocyte division and death rates from data, we use finite-dimensional approximations of the function spaces containing $\alpha$ and $\beta$ so that the problem is computationally tractable and theoretically sound [11, 12]. Previous work has established that division-linked changes in proliferation and death rates are an important aspect of an accurate mathematical model [42, 46]. One of the primary motivating assumptions behind the use of a PDE model for fitting CFSE data is that the dilution of CFSE dye by division allows for the structure variable (in this case, $y$) to be used as a surrogate for division number [13, 45, 47]. Thus division-dependent changes in proliferation and death rates are encapsulated in the structure dependence of the functions $\alpha$ and $\beta$.

While a straightforward implementation of structure dependence for $\alpha$ and $\beta$ has proven effective, the fact that the measured FI of a cell slowly decreases in time as a result of label loss indicates that one should take care in how the correlation between division number and structure variable is considered. As discussed at greater length in [6, 13], this label loss causes such correlation to lessen significantly. Alternatively, one might consider the total FI that *would* have been measured for a cell, if that cell did not experience any label loss. Mathematically, it is shown in [6] that this is equivalent to deriving a model in terms of an ideal label which does not decay and then changing one's frame of reference to a moving coordinate system in which the label does appear to decay (i.e., the one relative to which the data is actually taken). Similar situations frequently arise in mechanics and fluid dynamics, where discussions of Eulerian and Lagrangian formulations abound. The key argument is to identify a cell not by its current state $y$, but rather by the state it would have in the event it did not undergo label loss. For a cell with state $y$ at time $t$, this is equivalent to finding the intersection of the $y$-axis with the characteristic line passing through $(t, y)$. Given the characteristic lines

$$\frac{dy}{dt} = \frac{-ce^{-kt}}{\ln 10},$$

it follows that the cell located at $(t, y)$ was originally located (in the absence of division) at

$$s(t, y) = y + \frac{c}{k \ln 10}(1 - e^{-kt}).$$

It is shown in [6, 13] that the quantity $s(t, y)$ is more strongly correlated with division number than the quantity $y$. Moreover, the use of this 'translated coordinate' for the parameterization of $\alpha$ and $\beta$ provided a more accurate model of the observed data when compared to the simple implementation of spatial dependence. It was hypothesized that the improvement resulted from a more direct association between division number and proliferation rate. However, that analysis was done with a different label loss function, and we desire to repeat the analysis of [13] with the new model (19) and new label loss function (equivalent to (14)). Thus we consider four different parameterizations of the proliferation rate $\alpha$. In Section 5, results will be reported which demonstrate the effects of these different parameterizations on the effectiveness of the model.

First, we consider the simple case that $\alpha = \alpha(y)$. Given a fixed set of nodes $\{y_k\}$, we assume

$$\alpha = \alpha(y) = \sum_{k=1}^{K_\alpha} a_k l_k^{(\alpha)}(y), \tag{20}$$

where $l_k^\alpha(y)$ are piecewise linear spline functions satisfying

$$l_k^\alpha(y_j) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} .$$

It is assumed that $\alpha(0) = \alpha(3.5) = 0$. This assumption does not have a significant impact on the model as the nodes $\{y_k\}$ are chosen so that the proliferation rate can be varied as necessary at all values of the state variable where cells appear in the data. It does, however, add some measure of regularity to the computed proliferation rate function.

Alternatively, as discussed above, it may prove more accurate to use the translated coordinate $s$ in order to represent the proliferation rate of cells with a particular division number. Given a fixed set of nodes $\{s_k\}$, we assume

$$\alpha = \alpha(s) = \alpha(s(t, y)) = \sum_{k=1}^{K_\alpha} a_k l_k^{(\alpha)}(s), \tag{21}$$

where the functions $l_k^\alpha(s)$ are defined as above. Again, it is assumed that $\alpha(0) = \alpha(3.5) = 0$.

We also consider the possibility that the proliferation rate depends explicitly on time. Indeed, we see in the data in Figure 1 that there is no proliferation at least during the first 24 hours of the assay. However, by $t = 48$ hours, it is clear that the population has begun to divide. Thus the assumption of time dependence seems appropriate. As above, we can still consider the proliferation rate either in terms of the state variable $y$ or in terms of the translated coordinate $s$, in addition to its dependence on time. Given a set of nodes $\{y_k\}$ as above and a set of time nodes $\{t_m\}$, we parameterize

$$\alpha = \alpha(t, y) = \sum_{k=1}^{K_\alpha} \sum_{m=1}^{M} a_{km} l_k^{(\alpha)}(y) l_m^{(t)}(t), \tag{22}$$

where we now assume that the splines $l_k^{(\alpha)}(y)$ and $l_m^{(t)}(t)$ are piecewise linear in their respective variables. Again, we ensure smoothness in the forward simulation by requiring $\alpha(t, 0) = \alpha(t, 3.5) = 0$. It is also assumed that $\alpha(t, y) = 0$ for all $t \leq 24$ hours.

Finally, we also consider the case that $\alpha$ is parameterized in time as well as the translated coordinate $s$. Given nodes $\{t_m\}$ as above and nodes $\{s_k\}$ in the translated variable, the proliferation rate function is then

$$\alpha = \alpha(t, s) = \alpha(t, s(t, y)) = \sum_{k=1}^{K_\alpha} \sum_{m=1}^{M} a_{km} l_k^{(\alpha)}(s) l_m^{(t)}(t), \tag{23}$$

where we again assume the splines $l_k^{(\alpha)}(y)$ and $l_m^{(t)}(t)$ are piecewise linear. As before, it is assumed $\alpha(t, 0) = \alpha(t, 3.5) = 0$ and $\alpha(t, s) = 0$ for all $t \leq 24$ hours.

Results from [13, 45, 47] indicate that the death rate function need not be quite as complex as the proliferation rate function. After the first few generations, the death rate of cells seems to be roughly constant. There is little reason to suspect that the death rate function depends on time and we do not consider it here. As before, we consider using the state variable $y$ and the translated coordinate $s$ to parameterize the death rate function. Given nodes $\{y_k\}$ (which may be distinct from the nodes used in the estimation of the proliferation rate $\alpha$), we have

$$\beta = \beta(y) = \sum_{k=1}^{K_\beta} b_k l_k^{(\beta)}(y). \tag{24}$$

We assume $\beta(y) = b_1$ for all $y \in [0, y_1]$ and $\beta(y) = b_{K_\beta}$ for all $y \in [y_{K_\beta}, 3.5]$. Alternatively, using the translated coordinate, we have nodes $\{s_k\}$ and

$$\beta = \beta(s) = \beta(s(t, y)) = \sum_{k=1}^{K_\beta} b_k l_k^{(\beta)}(s) \tag{25}$$

with the assumptions $\beta(s) = b_1$ for all $s \in [0, s_1]$ and $\beta(s) = b_{K_\beta}$ for all $s \in [s_{K_\beta}, 3.5]$.

## 4.3 Computational Considerations

The change of variables $y = \log_{10}(x - x_a)$ is a parameter-dependent change of variables, technically requiring a 're-gridding' of the solution each time the parameters (specifically $x_a$) are changed in an optimization routine for the inverse problem. Because we use a time-stepping finite-difference method to compute the forward solution for a given set of parameters, this requirement does not constitute a great computational setback. In fact, observation of (19) reveals that the parameter $x_a$ does not appear directly in the equation to be solved, only in the change of variables that gives rise to the equation. The primary issues involve the determination of the initial condition from the data, and the comparison of the computed model solution to the data.

### 4.3.1 Formation of the Initial Condition

As discussed in Section 2, the cytometry data is reported as a time-series of histograms showing the numbers of cells counted into a given bin corresponding to a particular range of log CFSE FI values. That is, the histogram measures cells in the $z = \log_{10} x$ coordinate. While it is, in general, possible for the experimenter to set the bins to his/her own liking, we again recall that the current data set was obtained with bins already set. The original data set, as used in [13], was taken at 24 hour intervals over the course of 6 days. Data from Day 0 ($t = 0$ hours) is used to form the initial condition for the model in the manner described below; the remaining data is used to fit the model to the data. At time $t_i$, the data is stored as a set of ordered pairs $(z_i^j, n_i^j)$, $j = 1, \ldots, J(i)$ (the notation is meant to emphasize that the bins change each day). This ordered pair corresponds to the number of cells $n_i^j$ counted into the bin with left boundary $z_i^j$. (As a consequence, notice that we are unable to determine the width of the right-most bin at each time point; these points are simply removed from the data set).

Using the data taken at $t = 0$, we drew a smooth line through the data; ordered pairs representing the line were then determined using DataThief [63]. These smoothed histogram curves were then scaled upward into a smooth initial condition density $\hat{\Phi}(z)$ so that the total label content is the same for the smooth density as for the original histogram data. The results are depicted in Figure 6. Finally, given the initial condition $\hat{\Phi}(z)$, we transform this into an initial condition for $\tilde{n}(t, y)$ by noting that $y = \log_{10}(10^z - x_a)$ and using the label-preserving identity

$$\hat{\Phi}(z) = \frac{10^z}{10^z - x_a} \tilde{\Phi}(y(z)). \tag{26}$$

This, then, provides an initial condition for (19).

### 4.3.2 Comparison of the Model to the Data

While the procedure above provides a means of determining a smooth initial condition density $\tilde{\Phi}(y)$ from the smoothed histogram data, the comparison of the model, a density defined in terms of $y$, to the data, a histogram in terms of $z$, represents the opposite problem. In order to make this comparison, we need to perform two steps.

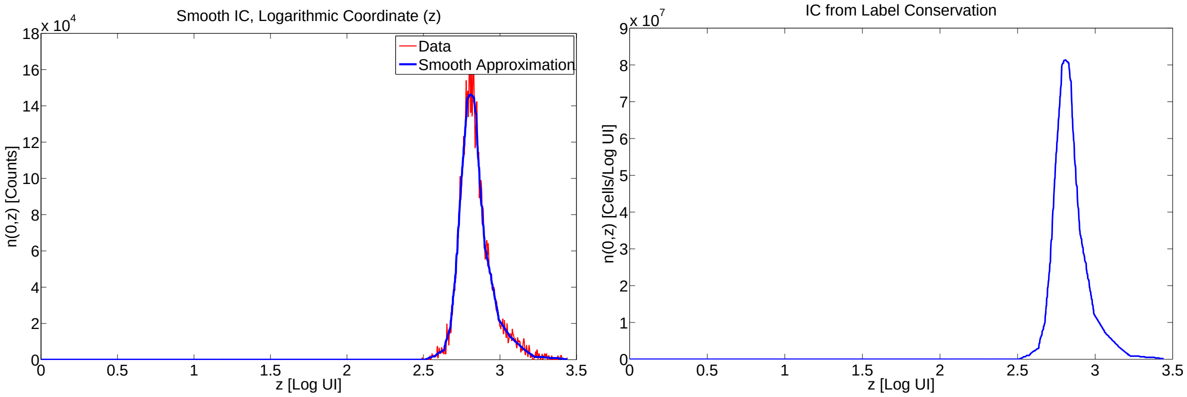

Figure 6: Left: Smoothed histogram data at $t = 0$ in the $z$ coordinate. Right: Density function computed from the smoothed histogram data.

First, we must transform the structured density $\tilde{n}(t, y)$ into a function of $z$. This is done analogously to the transformation (26) above. Second, we must transform this structured density into histogram numbers. To do this, we note that at time $t_i$, the total number of cells with FI between $z_i^j$ and $z_i^{j+1}$ is

$$I[\hat{n}](t_i, z_j) \equiv \int_{z_i^j}^{z_i^{j+1}} \hat{n}(t_i, z)dz \approx \left[\frac{\hat{n}(t_i, z_i^{j+1}) + \hat{n}(t_i, z_i^{j+1})}{2}\right]\left(z_i^{j+1} - z_i^j\right), \tag{27}$$

where the trapezoid rule has been used to approximate the integral. In general, we find that this is an effective method of obtaining histogram counts from the smooth density model solution. However, the varying sizes of the bins used to record the available data set poses somewhat of a problem. While the bins are generally regularly spaced, there are a few bins randomly placed in the data which are either much larger or much smaller than the neighboring bins. As a result, the histogram data $I[\hat{n}](t_i, z_j)$ computed from the smooth densities exhibits large jumps up or down at these points. This is strictly the result of the irregular bin sizes present in the data set and not the model solution itself. These jumps are problematic for the OLS procedure discussed below and as such these bins are removed from the data set. We emphasize again that, in future data sets, the bins can be set as needed, rather than being fixed in advance.

### 4.3.3 Finite Difference Computations

While (19) is defined on an infinite domain, all cells in the population maintain FI sufficiently greater than $x_a$, so that it is acceptable to solve (19) only on the domain $y \in [0, y_{\max}]$. In practice, we set $y_{\max}$ independent of the parameter $x_a$ and thus solve the equation (19) on the same computational domain regardless of the parameters (i.e., those passed in by the nonlinear optimization solver). Once the solution $n(t, y)$ is computed, it is possible to use (26) again to change variables back to $\hat{n}(t, z)$ for comparison to the data.

For the current data set, we use 512 evenly spaced nodes in the interval $y \in [0, 3.5]$. The forward solution is computed using a publicly available hyperbolic PDE solver written by L. Shampine which implements the Lax-Wendroff scheme.

## 4.4 Ordinary Least Squares Framework

Given the appropriate parameterizations of $\alpha$ and $\beta$, we now have a complete set of parameters $\theta = (x_a, c, k, \{a\}, \{b\})$ which define the model solutions. Thus in the analysis below we think of the parameterized model $\tilde{n}(t, y; \theta)$ satisfying (19) given parameter $\theta$. We now turn our attention to an inverse problem procedure which seeks to determine the parameters best describing the available data.

Following standard inverse problem procedure for ordinary least squares (OLS) [14, 20, 21], we assume that the data $\hat{n}_i^j$ represent an observation of the model solution evaluated at the true parameter $\theta_0$ with the addition

| Parameter | Minimum | Maximum | Units |
|---|---|---|---|
| $a_i$ | 0 | 1 | hr $^{-1}$ |
| $b_i$ | 0 | 1 | hr $^{-1}$ |
| $x_a$ | 0 | 100 | UI |
| $c$ | 0 | 0.1 | UI/hr |
| $k$ | 0 | 0.005 | hr $^{-1}$ |

Table 3: Summary of parameters $\theta = (x_a, c, k, \{a\}, \{b\})$ which define the model solution, with minimum values, maximum values, and units. Forward simulations of the model demonstrate the reasonableness of the bounds provided.

of some amount of noise. Thus, we can consider the data as a random variable

$$N_i^j = I[\hat{n}](t_i, z_j; \theta_0) + \mathcal{E}_{ij}, \tag{28}$$

where $\{\mathcal{E}_{ij}\}$ are random variables with $E[\mathcal{E}_{ij}] = 0$ and $Var(\mathcal{E}_{ij}) = \sigma^2$. We remark that the assumption of constant variance for the error terms is standard for OLS formulations of inverse problems. One can examine the accuracy of such an assumption ex post facto through the use of residual-based statistical tests [7, 14, 58]. In [13], such an analysis revealed that the actual error variance was neither constant nor proportional to the square of the model solution (a 'relative error model'). For the moment, we remark that the OLS assumption of constant variance, while possibly not exactly correct and hence not adequate for use in asymptotic parameter distributional analysis, is sufficient to provide a basis for computational parameter estimation, which will demonstrate the ability of the current model to fit the available data set. Given our uncertainty regarding the exact nature of the observed error process, we postpone a more detailed analysis of uncertainty in the estimated parameters (standard errors, confidence intervals, etc.) [7] for future work that will entail further experimental measurement error analysis and characterization.

Given the statistical model (28), we can write the data as realizations

$$n_i^j = I[\hat{n}](t_i, z_j; \theta_0) + \epsilon_{ij} \tag{29}$$

of the random variables (28). The goal of the OLS procedure is the determination of the parameter $\theta$ which minimizes the sum of squared residuals. Given the random variables $N_i^j$ from (28), the OLS estimator is

$$\theta_{\text{OLS}} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{I} \sum_{j=1}^{J(i)} (I[\hat{n}](t_i, z_j; \theta) - N_i^j)^2 = \arg\min J(\theta), \tag{30}$$

where $\Theta$ is a set of admissible parameters for the model (see Table 3). As the data $n_i^j$ are realizations of the random variables $N_i^j$, it follows that the OLS estimate

$$\hat{\theta}_{\text{OLS}} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{I} \sum_{j=1}^{J(i)} (I[\hat{n}](t_i, z_j; \theta) - n_i^j)^2 = \arg\min J(\theta), \tag{31}$$

is a realization of the OLS estimator $\theta_{\text{OLS}}$. This optimization was carried out with the MATLAB constrained optimization routine `fmincon`, which implements the BFGS algorithm at each step to solve a quadratic subproblem. Because such routines can become trapped in local minima, several initial iterates were tried for each optimization.

# 5 Results

The primary uncertainty in the inverse problem procedure is the choice of nodes $\{y_k\}$ for the estimation of the proliferation and death rates. We have no a priori information as to how many nodes should be used nor any information regarding where those nodes should be placed. To illustrate this point, in Figure 7 we depict the

| Division Number | $z$-axis Range | 7 Nodes | 13 Nodes | 25 Nodes |
|---|---|---|---|---|
| 6 | $[0.00, 1.05]$ | 0.0016 | 0.0015 | 0.0016 |
| 5 | $[1.05, 1.30]$ | 0.0043 | 0.0050 | 0.0052 |
| 4 | $[1.30, 1.60]$ | 0.0094 | 0.0103 | 0.0108 |
| 3 | $[1.60, 1.90]$ | 0.0166 | 0.0174 | 0.0206 |
| 2 | $[1.90, 2.25]$ | 0.0325 | 0.0308 | 0.0314 |
| 1 | $[2.25, 2.55]$ | 0.0284 | 0.0198 | 0.0266 |
| 0 | $[2.55, 3.50]$ | 0.0047 | 0.0097 | 0.0072 |

Table 4: Average proliferation rates (in units 1/hr) in terms of numbers of divisions undergone, computed from Figure 7. Using Figure 1, approximate ranges (in the coordinate $z$) corresponding to particular division numbers are determined. These are then used to compute the corresponding ranges in the variable $y$ using the estimated level of cellular autofluorescence. In spite of the differences in the numbers of nodes used in the parameterization of the proliferation rate function $\alpha(y)$, average proliferation values are estimated consistently for each generation.

estimated proliferation and death rate functions given three different choices of nodes $\{y_k\}$. First, seven nodes were evenly spaced in the interval $[1.125, 2.925]$. Next the number of nodes was increased to 13 so that the separation between nodes was halved (and so that the increase in parameters is a refinement to the model). This procedure was repeated and the estimation was performed a third time with 25 nodes. Intuitively, each refinement must provide a more accurate fit of the model to the data, as the total OLS cost cannot increase as the number of parameters is increased. However, we also see in Figure 7 that as the number of nodes increases, the estimated proliferation rate function becomes less regular. Conversely, we see that some measure of regularity can be imposed on the function $\alpha(y)$ by choosing the proper set of nodes with which to estimate it (a so-called 'regularization by discretization' [10, 11]). While we do have available residual-sum-of-squares based statistical tests [7, 8, 14] to quantify the improvement in the model with each refinement, we choose to balance this additional information with a desire to estimate a semi-regular function $\alpha$.

It is worth remarking further that the increasingly complex structure of the function $\alpha(y)$ (as the number of nodes is increased) is only a relic of the estimation procedure and has nothing to do with any meaningful information regarding the population of cells being studied. In order to verify this, in Table 4 we present, for each parameterization of the function $\alpha(y)$ discussed in the previous paragraph, the average value of the estimated proliferation rate in terms of the numbers of divisions the cells have undergone. To compute these values, Figure 1 is used to determine approximate ranges (in the coordinate $z$) corresponding to each generation of cells. By changing these ranges from the variable $z$ to $y$ (in which the estimation of the proliferation rate function was performed), the average value of the proliferation rate function can be determined in each range. As seen in Table 4, the estimates are reasonably consistent regardless of the number of nodes used in the estimation.

Given the above discussion, we choose 13 nodes for the estimation of the proliferation rate function $\alpha(y)$ and 5 nodes for the estimation of the death rate function $\beta(y)$ in an effort maximize the flexibility of the estimation while also maintaining some regularity in the estimated functions. The OLS best-fit solution is shown in Figure 8. The optimal proliferation and death rate functions are shown graphically in the center panels of Figure 7, with numerical values provided in Tables 5 and 6. The total OLS cost for the estimation is $J(\hat{\theta}_{\text{OLS}}) = 1.7270 \times 10^{12}$, with $x_a = 8.2316$, $c = 5.6169 \times 10^{-3}$, and $k = 1.1203 \times 10^{-8}$.

We observe that, while the OLS best-fit solution for time-independent proliferation is accurate for $t = 96$ and 120 hours, the model predicts far too many cells with high generation number at $t = 24$ and 48 hours. This seems to be a manifestation of the absence of a delay (in the form of time dependence) between the time cells are stimulated and the time at which those stimulated cells divide. Thus, in addition to the arguments of Section 4.2, we have a mathematical rationale for the incorporation of time dependence into the proliferation rate. While the model does not accurately count the numbers of cells in the earlier time points, we do remark that the Gompertz label loss model and the incorporation of cellular AutoFI do accurately predict the *location* (along the horizontal axis) of the subsequent generations of cells in culture. Thus, it appears safe to conclude that the parameter $\gamma$ from [13, 47] has been effectively removed from any modeling needs.

Given this discussion, we next consider the possibility that the proliferation rate function $\alpha(t, y)$ may depend on time as well as on the structure variable $y$ (see Section 4.2) in order to better estimate the numbers of cells in each generation at a given time. We would like to make use of model refinement techniques in order to quantify

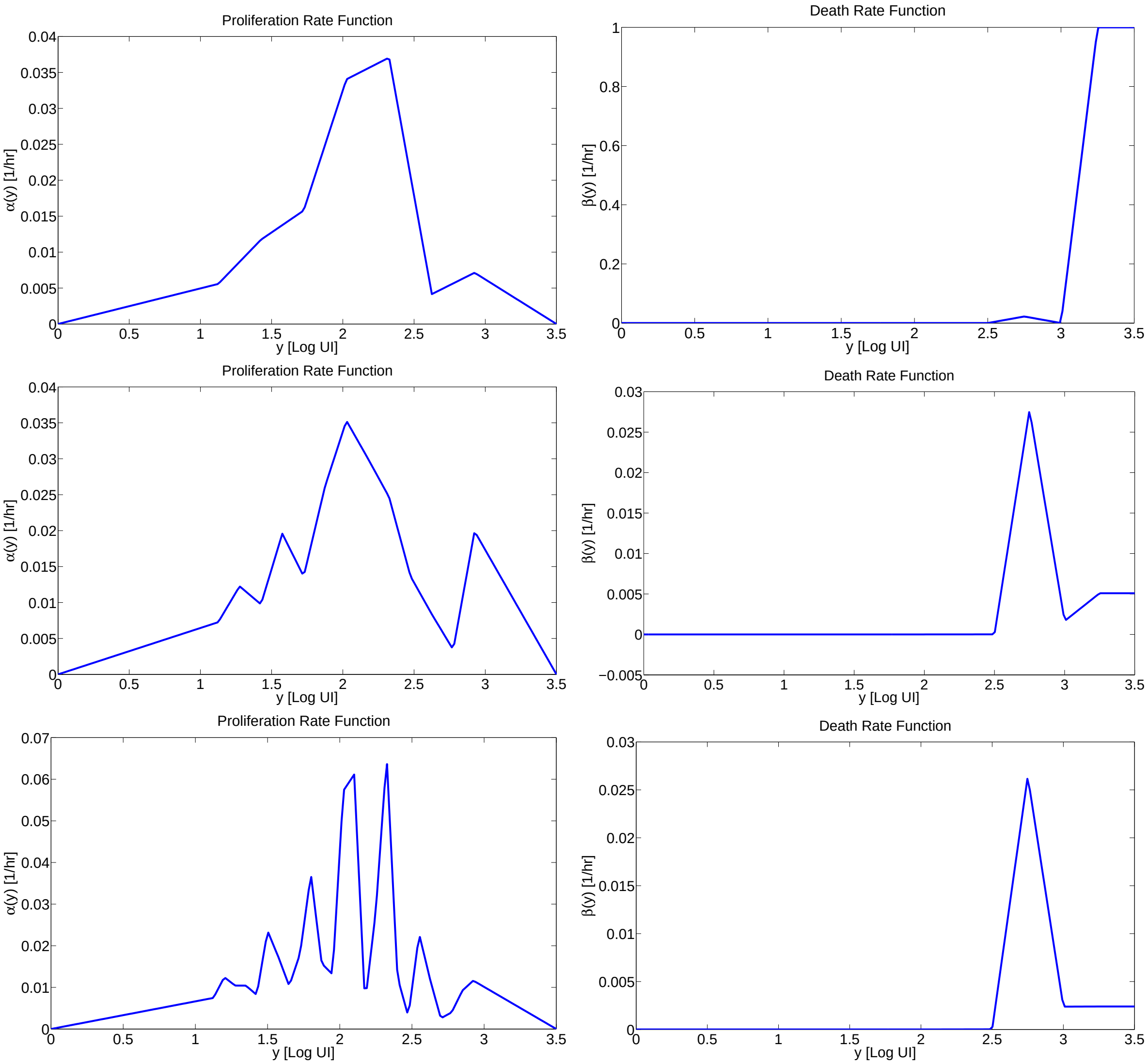

Figure 7: Left: Estimated proliferation rate function for three difference choices of nodes. Top: 7 nodes evenly spaced in [1.125,2.925]. Middle: 13 nodes evenly spaced in [1.125,2.925]. Bottom: 25 nodes evenly spaced in [1.125,2.925]. Note that, while the overall shape of $\alpha(y)$ remains largely the same, the middle figure seems to provide the most information while remaining some semblance of regularity. These functions can be used to determine the average rate of proliferation in terms of the number of divisions undergone (Table 4). Right: the corresponding estimated death rate function $\beta(y)$, estimated using 5 fixed nodes in each case.

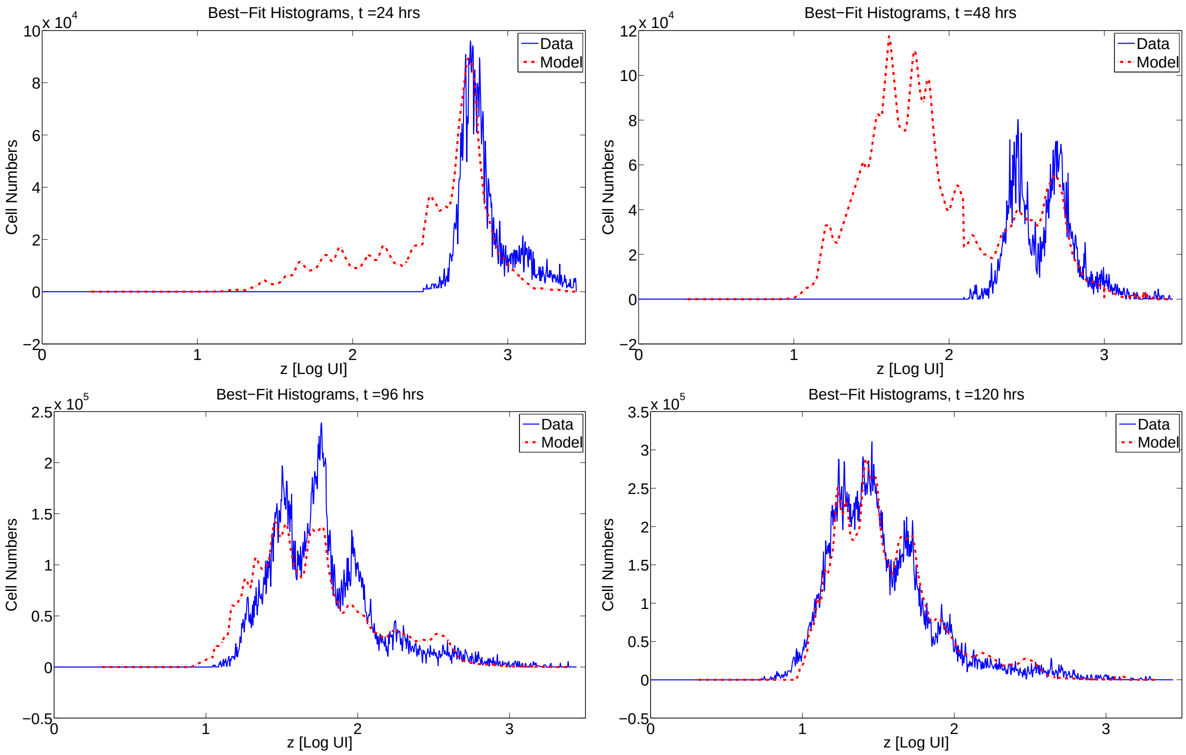

Figure 8: OLS best-fit solution with $\alpha = \alpha(y)$ (13 nodes), $\beta = \beta(y)$ (5 nodes). 21 total parameters in the model, total cost $J(\hat{\theta}_{\text{OLS}}) = 1.7270 \times 10^{12}$. While the model clearly is not accurate in allowing far too many cells with large division number too early in time, the *locations* of the division peaks along the horizontal axis are quite accurate, in support of the role of autofluorescence as well as the Gompertz decay of label.

| $y_k^{(\alpha)}$ | $a_k$ |
|---|---|
| 1.1250 | 0.0073 |
| 1.2750 | 0.0123 |
| 1.4250 | 0.0097 |
| 1.5750 | 0.0196 |
| 1.7250 | 0.0136 |
| 1.8750 | 0.0262 |
| 2.0250 | 0.0353 |
| 2.1750 | 0.0301 |
| 2.3250 | 0.0247 |
| 2.4750 | 0.0137 |
| 2.6250 | 0.0084 |
| 2.7750 | 0.0034 |
| 2.9250 | 0.0199 |

Table 5: Results for the OLS estimation of $\alpha(y)$ with 13 nodes. This estimated proliferation rate function is shown graphically in the left-center panel of Figure 7. Average rate of division in terms of division number is computed in Table 4.

| $y_k^{(\beta)}$ | $b_k$ |
|---|---|
| 2.0000 | 0.0000 |
| 2.5000 | 0.0000 |
| 2.7500 | 0.0278 |
| 3.0000 | 0.0017 |
| 3.2500 | 0.0051 |

Table 6: Results for the OLS estimation of $\beta(y)$ with 5 nodes when $\alpha = \alpha(y)$ is estimated with 13 nodes. This estimated death rate function is shown graphically in the right-center panel of Figure 7.

the resulting improvement in the fit of the model to data while also accounting for the increased complexity of the model. Thus, we use the same 13 nodes as above for the structural discretization of $\alpha$. For the time discretization, nodes $\{t_m\} = [48, 60, 72, 96, 120]$ are used. The death rate function $\beta(y)$ is estimated exactly as before. This parameterization results in a model with 73 parameters. After calibration to the data, the resulting cost is $J(\hat{\theta}_{\text{OLS}}) = 3.1302 \times 10^{11}$ with $x_a = 6.2698$, $c = 4.8123 \times 10^{-3}$, and $k = 9.8091 \times 10^{-8}$; the fit of the model to the data is shown in Figure 9. It is clear from the figure that the improvement in fitting the model to the data is quite significant. Moreover, because the inclusion of time-dependence is a refinement of the time-independent model, residual-sum-of-squares-based statistical tests exist to quantify whether the increase in complexity of the model (from 21 to 73 parameters) is justified by the resulting reduction in cost. Using the method described in [14, Ch. 3], we find that the time-independent model can be rejected in favor of time-dependent proliferation with very high ($> 99.999\%$) confidence.

Thus we see that, once the proliferation rate is allowed to vary as a function of time, the model very closely mimics the data in terms of the numbers of cells in each generation at a given time. Moreover, we again point out that the physiological explanation for the the dilution of FI by division, as well as the Gompertz model for natural FI decay do an excellent job of predicting the locations (along the horizontal axis) of each generation of cells. Still, we continue further to consider one more potential improvement to the model. Following the analysis of [13] and the discussion of Section 4.2, we consider parameterizing the proliferation and death rate functions in terms of the 'translated coordinate' $s$. As discussed previously, it is expected that this coordinate correlates much more closely with division number than the coordinates $z$ or $y$. As such, estimation of the proliferation and death rates in terms of this quantity should provide a more meaningful (and less biased) estimate when these estimated functions are analyzed in terms of division number (in the manner of Table 4). It is worth noting the parameterization of the functions $\alpha$ and $\beta$ in terms of $s$ is not a model refinement (compared to parameterization in terms of $y$) so that the model comparison tests described above are not directly applicable. While additional (e.g. information-theoretic) tests could be used, we forgo that analysis here in the interest of brevity. However, as will be shown, parameterization in terms of $s$ does in fact provide a more meaningful correlation between the estimated cell turnover rates and division number (in additional to providing a slightly lower cost!), which justifies its use.

The nodes used from the proliferation and death rate functions, as well as the estimated rates at those nodes, are given in Tables 7 and 8, respectively, and the functions are shown graphically in Figure 10. As before, 73 parameters arise in this parameterization of the model. The total cost is $J(\hat{\theta}_{\text{OLS}}) = 3.0901 \times 10^{11}$, with $x_a = 6.4053$, $c = 5.5246 \times 10^{-3}$, and $k = 5.0323 \times 10^{-4}$; the fit of the model to the data is shown in Figure 11.

Visually, the fit of this model (using $s$ for the structure discretization of the proliferation and death rates) is comparable to the previous model (using $y$), and the cost is slightly lower. The significant advantage in using $s$, as noted above, is that the translated coordinate $s$ is more strongly correlated with division number. To see this, the data from Figure 1 are shown in the translated coordinate in Figure 12. While there is still some overlap among the generations of cells in the histogram data, the translated coordinate provides an axis on which cells do not drift as they slowly lose CFSE FI. Particularly when compared to Figure 1, we see that it is much easier to assign distinct regions of the $s$ axis to particular division numbers when compared to using the $z$ (and thus $y$) axis. Moreover, because cells do not drift to the left on the $s$ axis, regions assigned to particular division numbers remain valid for all time.

Given the near-alignment of the generations of cells in the translated coordinate $s$, a similar analysis to that presented in Table 4 can be performed. By determining intervals (in the $s$ coordinate) corresponding to particular

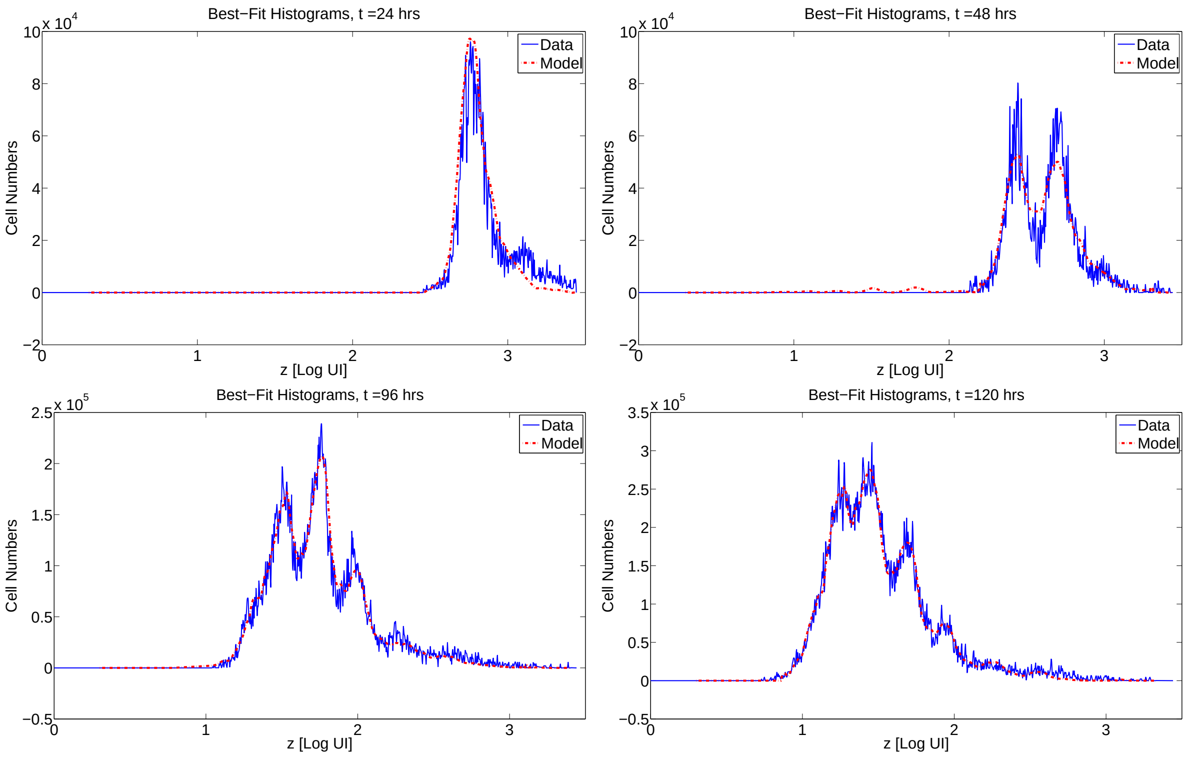

Figure 9: OLS best-fit solution with $\alpha = \alpha(t, y)$, $\beta = \beta(y)$. 73 total parameters in the model, total cost $J(\hat{\theta}_{\text{OLS}}) = 3.1302 \times 10^{11}$.

| | $t_k$ | | | | |
|---|---|---|---|---|---|
| $s_k^{(\alpha)}$ | 48 | 60 | 72 | 96 | 120 |
| 1.1875 | 0.0713 | 0.1404 | 0.0000 | 0.0000 | 0.0167 |
| 1.3375 | 0.2028 | 0.0522 | 0.0001 | 0.0000 | 0.0175 |
| 1.4875 | 0.6036 | 0.0303 | 0.0376 | 0.0009 | 0.0281 |
| 1.6375 | 0.2896 | 0.0138 | 0.0004 | 0.0075 | 0.0251 |
| 1.7875 | 0.0618 | 0.0001 | 0.0000 | 0.0409 | 0.0220 |
| 1.9375 | 0.0091 | 0.0345 | 0.0020 | 0.0119 | 0.0220 |
| 2.0875 | 0.0837 | 0.0002 | 0.0400 | 0.0326 | 0.0391 |
| 2.2375 | 0.0018 | 0.1956 | 0.0083 | 0.0001 | 0.0231 |
| 2.3875 | 0.0050 | 0.0059 | 0.0962 | 0.0394 | 0.0463 |
| 2.5375 | 0.0000 | 0.1949 | 0.0128 | 0.0050 | 0.0000 |
| 2.6875 | 0.0155 | 0.1101 | 0.1528 | 0.0422 | 0.0239 |
| 2.8375 | 0.0351 | 0.0446 | 0.0000 | 0.0000 | 0.0000 |
| 2.9875 | 0.0346 | 0.0000 | 0.0012 | 0.1317 | 0.0092 |

Table 7: Results for the OLS estimation of $\alpha(t, s)$. This estimated proliferation rate function is shown graphically in Figure 10. Average rate of division in terms of division number is computed in Table 13.

| $s_k^{(\beta)}$ | $b_k$ |
|---|---|
| 2.0000 | 0.0054 |
| 2.5000 | 0.0000 |
| 2.7500 | 0.0238 |
| 3.0000 | 0.0061 |
| 3.2500 | 0.0000 |

Table 8: Results for the OLS estimation of $\beta(s)$ when $\alpha = \alpha(t, s)$. This estimated death rate function is shown graphically in Figure 10

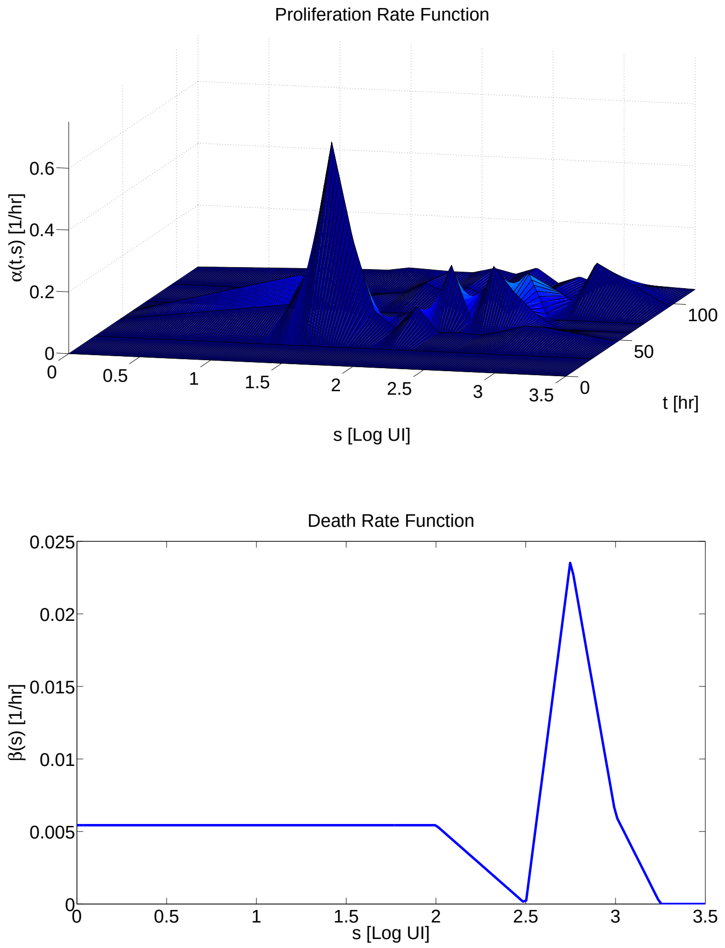

Figure 10: OLS best-fit proliferation and death rate function $\alpha(s, t)$ (top) and $\beta(s)$ (bottom), respectively, when the proliferation rate is assumed to depend on both $s$ and $t$.

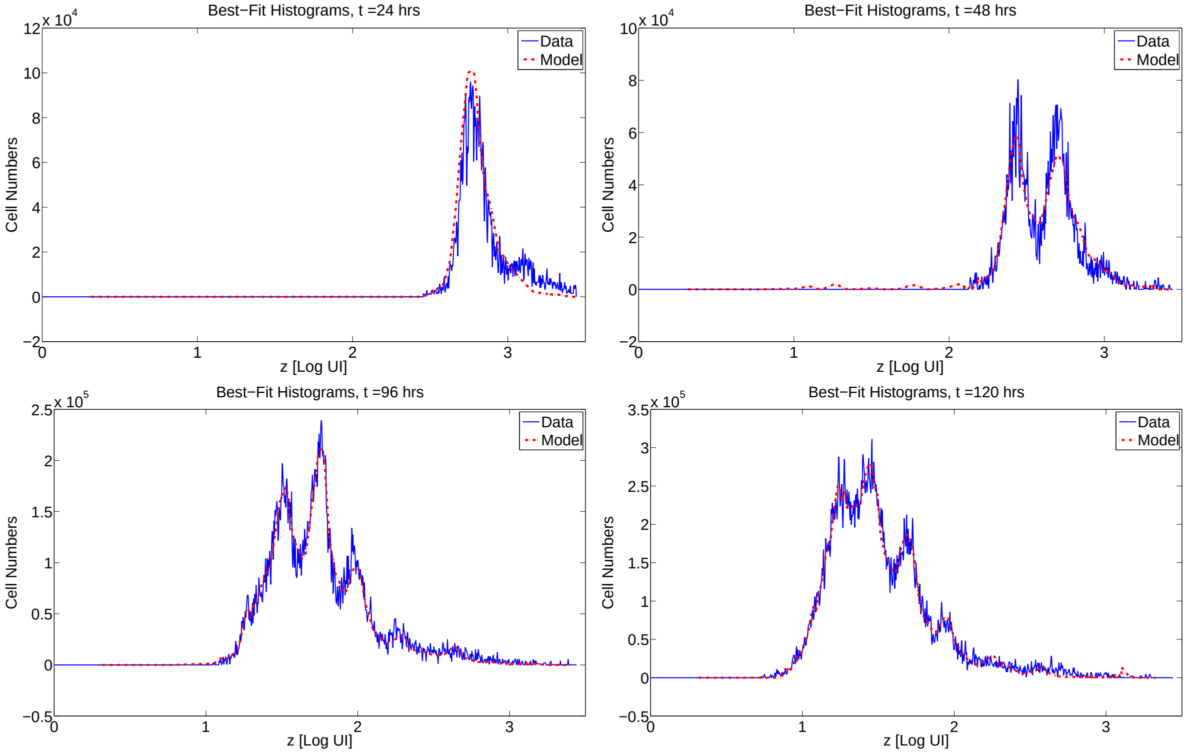

Figure 11: OLS best-fit solution with $\alpha = \alpha(t,s)$, $\beta = \beta(s)$. 73 total parameters in the model, total cost $J(\hat{\theta}_{\mathrm{OLS}}) = 3.0901 \times 10^{11}$.

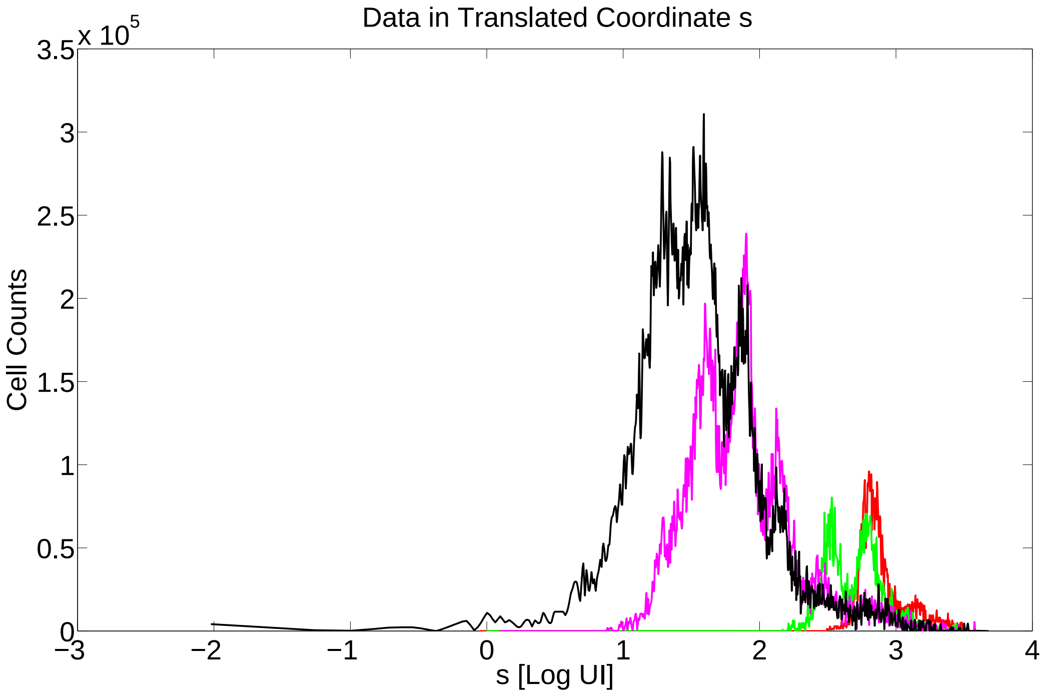

Figure 12: Data for $t = 24, 48, 96, 120$, respectively, plotted in the translated coordinate $s$. Observe that the peaks corresponding to distinct division numbers closely align.

division numbers, the average rate of proliferation for cells having undergone a specified number of divisions can be computed. Because we now have a proliferation rate which depends explicitly on time, we compute the average proliferation rate (in terms of the number of divisions undergone) and display this information as a function of time (rather than averaging in time as well) in Figure 13, thus preserving what we believe to be an important feature of the population of cells (that the proliferation rates change in time).

# 6 Discussion

In this document, we have presented significant modifications and clarifications to the results of [13] and [47]. Of primary importance is that the parameter $\gamma$, used in the previous model to heuristically explain the dilution of CFSE resulting from cell division, has been replaced by a physiologically-based mechanism which accounts for cellular autofluorescence. Also important is the use of the Gompertz decay process to explain the natural decay of CFSE observed in the data. We have seen that these two improvements in the model are fully capable of fitting the same data as in [13]. Moreover, these revisions provide clarity to the model because they can be understood in terms of physiologically relevant and easily observable features of the data.

We also revise the manner in which the model, a structured density, is compared to the data, a series of histograms. Because the data used in this report is already in histogram form, the irregular size of certain bins caused some computational difficulty. In future experiments, direct control over the histogram bin spacing will remove these difficulties. We also hope to use future data sets to determine an accurate statistical model for the error/noise in a given data set, and how that model changes as the histogram bins are selected in different ways. As the ultimate goal of any model of the immune response is the comparison of changing intra- and extracellular conditions on proliferative behavior [30, 32], uncertainty quantification in the form of confidence intervals are necessary to facilitate such a comparison. Asymptotic theory for sum-of-squares-based estimators and model comparison tests [7, 8, 20, 58] exist, but rely upon a correct underlying statistical model assumption for the data. While we assume a constant variance model here (28), it has been shown that this is not correct [13], and thus any confidence intervals computed from such a formulation would be in error. (This is not to say that the estimated parameters reported here are invalid, but only that we cannot continue further to quantify the certainty of those estimates without additional information.) Determination of an accurate statistical model is of vital importance for the unbiased estimation of standard errors and confidence intervals for the estimated model parameters [14].

Similarly, the mechanism responsible for the apparently spurious measurement at $t = 72$ hours must be properly accounted for in future work. As has been shown, the total mass of CFSE in the cell culture is measured to increase from $t = 48$ hours to $t = 72$ hours in [13]. The inability of the current model to describe this behavior is not directly a shortcoming of the model itself (as any method, e.g., deconvolution with a series of gaussian curves, would suffer from a similarly biased result) but rather represents an incomplete understanding of the nature of the observation process. Indeed, it is an asset of the current model, derived from conservation principles, that such a feature has been noticed. Thus, in addition to the statistical model discussed in the previous paragraph, future work will need to focus on establishing an observation model to accurately account for the manner in which the cell population data is represented.

In support of the results of [13], parameterization of the proliferation and death rates $\alpha$ and $\beta$ in terms of the translated or moving variable $s$ provides an improvement to the OLS fit of the model to the data. Beyond this minor improvement, the introduction of this variable was motivated by the fact that, when all data is placed in the coordinate $s$ (Figure 12), the peaks corresponding to distinct division numbers align much more closely than when presented in terms of the measurement variable $z$ (Figure 1). As such, the estimation of the proliferation and death rate functions in terms of this variable should more directly relate division number to the rate at which cells in a given generation divide. It follows that average rates of proliferation can be calculated in terms of the number of divisions a cell has undergone, and the dependence of these rates on time can also be explored (Table 13). Thus, we find that the translated coordinate $s$, as a result of its strong correlation with division number, permits the estimated functions $\alpha$ and $\beta$ to be easily related to intuitive measures of a lymphocyte response such as mean division time, mean doubling time, etc. Such information provides a nearly complete quantitative picture of a dynamic T cell responsiveness and thereby may be helpful for mechanistic studies of immune control.

The mathematical model presented here is more complex than many of the existing frameworks for understanding cell turnover kinetics. While the number of parameters will vary depending upon the manner in which nodes for the estimation of the proliferation and death rate functions $\alpha$ and $\beta$ are chosen, the best-fit results

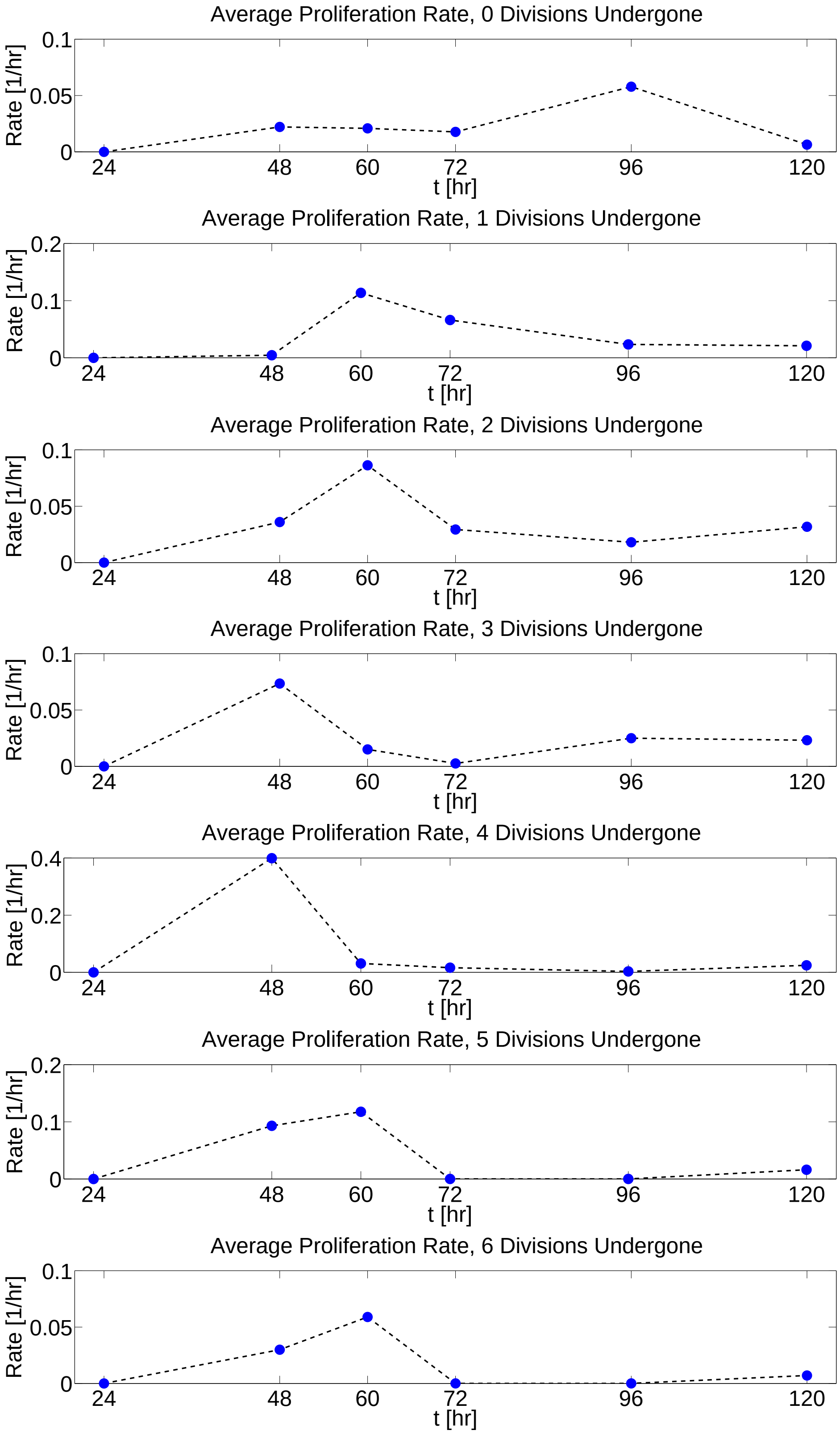

Figure 13: Average proliferation rate as a function of time for each generation of cells.

presented in this report require 73 parameters. Optimization times range from 1 to 8 hours, depending upon the accuracy of the initial iterate and the tolerances selected for the optimization routine. In spite of this additional complexity, the current model accurately predicts CFSE-based proliferation dynamics and does so by directly addressing histogram data from the assay. By avoiding the need for any deconvolution techniques to extract cell numbers (per generation) from the histograms, some potential bias and/or error is avoided. Additionally, by directly addressing quantities such as autofluorescence and the natural decay of label, their effects on the observed behavior of the population can be quantified. For instance, we might generalize the model presented here to account for AutoFI which changes as a function of time (as cells are activated and/or enter a quiescent state) or which varies from cell to cell in the population. While the exact shape of the estimated proliferation and death rate functions may change with various choices of nodes (Figure 7), we have seen that, for reasonable parameterization, the average proliferation rates (as a function of division number) are consistently estimated (Table 4). When estimating the time-dependent proliferation rate, it should be noted that, for high division number, the rate estimated for early times must be interpreted with caution: the rate is ultimately meaningless until cells have emerged in the population which divide at that rate. One potential solution to this caveat is to use a more complex (e.g. non-rectangular) grid for the estimation of $\alpha(t,s)$.

In the current model formulation, the rates of proliferation and/or death are essentially exponential (in the sense that the rate of change of the population, $\frac{\partial n}{\partial t}$ is directly proportional to the total population $n(t,y)$) with rates $\alpha$ and $\beta$, respectively. We do not need to make use of delays or minimum cell cycle times in order to fit the data (as the time-dependence of the function $\alpha$ accomplishes the same effect). An interesting generalization of the current model would be the incorporation of volume structure. As cells would need to progress from size $V$ to $2V$ before division, this would naturally require the incorporation of some cell cycle time. Moreover, forward scatter (FSC) of laser light may possibly be used as an observable surrogate for cell size. Importantly, the current model assumes that all cells in the population behave in exactly the same manner. However, recent work has demonstrated that there are cohorts of closely related cells whose behavior is correlated in some way [25, 35, 64, 67]. Thus it may be necessary to examine in the framework of [4, 5, 9] the effects of probabilistically distributed parameters within the population. This same framework could be used to describe subpopulations of cells with varying levels of AutoFI (as discussed in the previous paragraph) or which internally process/catabolize intracellular die at different rates.

Using the data set from [13, 47], we have shown that the incorporation of autofluorescence and Gompertz decay of label provide a mathematical model with firm physiological underpinnings which can accurately describe CFSE histogram data directly. Because of the nonparametric manner in which the proliferation and death rates are estimated, this model is able to encapsulate a wide variety of proliferative responses as various types of cells are subjected to a variety of experimental conditions and then measured. We are actively working to collect additional data sets with which to demonstrate the widespread applicability of this model, as well as to use this model in a systematic fashion to analyze how the estimated parameters vary under changing experimental and biological conditions. As more information becomes available regarding the complex processes involved in cell proliferation, we are confident that the model discussed here provides a firm physiological foundation upon which CFSE-based assay data can be understood. We strongly believe that the ideas and results presented here will form an important interpretive framework with a wide array of applications in experimental settings, diagnostic tests [28], and perhaps in a more integrated model of cell dynamics [39, 44]. The authors are also grateful to referees for constructive comments and suggestions for improvements in this manuscript.

## Acknowledgments:

This research was supported in part by the National Institute of Allergy and Infectious Disease under grant NIAID 9R01AI071915, in part by the U.S. Air Force Office of Scientific Research under grant AFOSR-FA9550-09-1-0226, in part by the Russian Foundation for Basic Research (Grant 11-01-00117), in part by the Program of the Russian Academy of Sciences Basic Research for Medicine, in part by the Agence Nationale de la Recherche, Grant No. ANR-09-BLAN-0218 TOPPAZ, in part by the Deutsche Forschungsgemeinschaft and in part by grant SAF2010-21336 from the Spanish Ministry of Science and Innovation. The authors are also grateful to referees for constructive comments and suggestions for improvements in this manuscript.

# References

[1] O. Arino, E. Sanchez, and G.F. Webb, Necessary and sufficient conditions for asynchronous exponential growth in age structured cell populations with quiescence, *J. Mathematical Analysis and Applications*, **215** (1977), 499–513.

[2] B. Asquith, C. Debacq, A. Florins, N. Gillet, T. Sanchez-Alcaraz, A. Mosley, and L. Willems, Quantifying lymphocyte kinetics in vivo using carboxyfluorein diacetate succinimidyl ester, *Proc. R. Soc. B* **273** (2006), 1165–1171.

[3] J.E. Aubin, Autoflouresecence of viable cultured mammalian cells, *J. Histochem. Cytochem.,* **27** (1979), 36–43.

[4] H.T. Banks, L.W. Botsford, F. Kappel, and C. Wang, Modeling and estimation in size structured population models, LCDS/CSS Report 87-13, Brown University March, 1987; *Proc. 2nd Course on Math. Ecology* (Trieste, December 8-12, 1986) World Scientific Press, Singapore, 1988, 521–541.

[5] H.T. Banks and J.L. Davis, A comparison of approximation methods for the estimation of probability distributions on parameters, *Appl. Num. Math.* **57** (2007), 753–777.

[6] H.T. Banks, Frederique Charles, Marie Doumic, Karyn L. Sutton, and W. Clayton Thompson, Label structured cell proliferation models, *Appl. Math. Letters* **23** (2010), 1412–1415; doi:10.1016/j.aml.2010.07.009

[7] H.T. Banks, M. Davidian, J. Samuels, and K.L. Sutton, An inverse problem statistical methodology summary, CRSC-TR08-01, NCSU, January, 2008; Chapter 11 in *Mathematical and Statistical Estimation Approaches in Epidemiology*, G. Chowell, et al., eds., Berlin Heidelberg New York, 2009, pp. 249–302.

[8] H.T. Banks and B.G. Fitzpatrick, Inverse problems for distributed systems: statistical tests and ANOVA, LCDS/CSS Report 88-16, Brown University, July 1988; *Proc. International Symposium on Math. Approaches to Envir. and Ecol. Problems*, Springer Lecture Notes in Biomath., **81** (1989), 262–273.

[9] H.T. Banks and B.F. Fitzpatrick, Estimation of growth rate distributions in size-structured population models, CAMS Tech. Rep. 90-2, Univ. of Southern California, January, 1990; *Quart. Appl. Math.* **49** (1991), 215–235.

[10] H.T. Banks and D.W. Iles, On compactness of admissible parameter sets: convergence and stability in inverse problems for distributed parameter systems, ICASE Report 86-38, NASA Langley Res. Ctr., Hampton, Virginia, 1986; *Proc. Conf. on Control Systems Governed by PDE's*, Gainesville, Florida. Springer Lecture Notes in Control and Inf. Sci., **97** (1987), 130–142.

[11] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems* Birkhauser, Boston, 1989.

[12] H.T. Banks and M. Pedersen, Well-posedness of inverse problems for systems with time dependent parameters, CRSC-TR08-10, NCSU, August 2008; *Arab. J. Sci. Eng. Math.* **1** (2009), 39–58.

[13] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, Gennady Bocharov, Dirk Roose, Tim Schenkel, and Andreas Meyerhans, Estimation of cell proliferation dynamics using CFSE data, CRSC-TR09-17, NCSU, August, 2009; *Bull. Math. Biol.* **70** (2011), 116–150; doi:10.1007/s11538-010-9524-5.

[14] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton London New York, 2009.

[15] B. Basse, B. Baguley, E. Marshall, G. Wake, D. Wall, Modelling the flow cytometric data obtained from unperturbed human tumour cell lines: Parameter fitting and comparison, *Bull. Math. Biol.*, **67** (2005), 815–830.

[16] F. Bekkal Brikci, J. Clairambault, B. Ribba, and B. Perthame, An age-and-cyclin-structured cell population model for healthy and tumoral tissues, *J. Math. Biol.* **57** (2008), 91–110.

[17] G. Bell and E. Anderson, Cell Growth and Division I. A Mathematical Model with Applications to Cell Volume Distributions in Mammalian Suspension Cultures, *Biophysical Journal* **7** (1967), 329–351.

[18] S. Bonhoeffer, H. Mohri, D. Ho, and A.S. Perelson, Quantification of cell turnover kinetics using 5-Bromo-2'-deoxyuridine, *J. Immunology* **64** (2000), 5049–5054.

[19] R. Callard and P.D. Hodgkin, Modeling T- and B-cell growth and differentiation, *Immunological Reviews* **216** (2007), 119–129.

[20] R.J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman Hall, London, 2000.

[21] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London, 2000.

[22] R.J. DeBoer, V.V. Ganusov, D. Milutinovic, P.D. Hodgkin, and A.S. Perelson, Estimaing lymphocyte division and death rates from CFSE data, *Bull. Math. Biol.* **68** (2006), 1011–1031.

[23] R.J. DeBoer and Alan S. Perelson, Estimating division and death rates from CFSE data, *J. Comp. and Appl. Mathematics* **184** (2005), 140–164.

[24] E.K. Deenick, A.V. Gett, P.D. Hodgkin, Stochastic model of T cell proliferation: a calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival, *J. Immunology* **170** (2003), 4963–4972.

[25] K. Duffy and V. Subramanian, On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics, *J. Math. Biol.* **59** (2009), 255–285.

[26] J.Z. Farkas, Stability conditions for the non-linear McKendrick equations, *Appl. Math. and Comp.* **156** (2004), 771–777.

[27] J.Z. Farkas, Stability conditions for a non-linear size-structured model, *Nonlinear Analysis: Real World Applications* **6** (2005), 962–969.

[28] D.A. Fulcher and S.W.J. Wong, Carboxyfluorescein diacetate succinimidyl ester-based assays for assessment of T cell function in the diagnostic laboratory, *Immunology and Cell Biology* **77** (1999) 559–564.

[29] V.V. Ganusov, S.S. Pilyugin, R.J. De Boer, K. Murali-Krishna, R. Ahmed, and R. Antia, Quantifying cell turnover using CFSE data, *J. Immunological Methods* **298** (2005), 183–200.

[30] A.V. Gett and P.D. Hodgkin, A cellular calculus for signal integration by T cells, *Nature Immunology* **1** (2000), 239–244.

[31] B.F. de St. Groth, A.L. Smith, W. Koh, L. Girgis, M. Cook, P. Bertolino, Carboxyfluorescein diacetate succinimidyl ester and the virgin lymphocyte: a marriage made in heaven, *Immunology and Cell Biology* **77** (1999) 530–538.

[32] J. Hasbold, A.V. Gett, J.S. Rush, E. Deenick, D. Avery, J. Jun, and P.D. Hodgkin, Quantitative analysis of lymphocyte proliferation and differentiation in vitro using carboxyfluorescein diacetate succinimidyl ester, *Immunology and Cell Biology* **77** (1999) 516–522.

[33] E.D. Hawkins, Mirja Hommel, M.L Turner, Francis Battye, J Markham and P.D Hodgkin, Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data, *Nature Protocols* **2** (2007), 2057–2067.

[34] E.D. Hawkins, M.L. Turner, M.R. Dowling, C. van Gend, and P.D. Hodgkin, A model of immune regulation as a consequence of randomized lymphocyte division and death times, *Proc. Natl. Acad. Sci* **104** (2007), 5032–5037.

[35] E.D. Hawkins, J.F. Markham, L.P. McGuinness, and P.D. Hodgkin, A single-cell pedigree analysis of alternative stochastic lymphocyte fates, *Proc. Natl. Acad. Sci* **106** (2009), 13457–13462.

[36] P.D. Hodgkin, J. Lee, A.B. Lyons, B cell differentiation and isotype switching is related to division cycle number, *J. Exp. Med.* **184** (1996), 277–281.

[37] O. Hyrien and M.S. Zand, A mixture model with dependent observations for the analysis of CFSE-labeling experiments, *J. American Statistical Association* **103** (2008), 222–239.

[38] O. Hyrien, R. Chen, and M.S. Zand, An age-dependent branching process model for the analysis of CFSE-labeling experiments, *Biology Direct* **5** (2010), Published Online.

[39] D.E. Kirschner, S.T. Chang, T.W. Riggs, N. Perry, and J.J. Linderman, Toward a multiscale model of antigen presentation in immunity, *Immunological Reviews* **216** (2007), 93–118.

[40] M. Kot, *Elements of Mathematical Ecology*, Cambridge UP: Cambridge, UK, 2001.

[41] H.Y. Lee, E.D. Hawkins, M.S. Zand, T. Mosmann, H. Wu, P.D. Hodgkin, and A.S. Perelson, Interpreting CFSE obtained division histories of B cells in vitro with Smith-Martin and Cyton type models, *Bull. Math. Biol.* **71** (2009), 1649–1670.

[42] H.Y. Lee and A.S. Perelson, Modeling T cell proliferation and death in vitro based on labeling data: generalizations of the Smith-Martin cell cycle model, *Bull. Math. Biol.* **70** (2008), 21–44.

[43] K. Leon, J. Faro, and J. Carneiro, A general mathematical framework to model generation structure in a population of asynchronously dividing cells, *J. Theoretical Biology* **229** (2004), 455–476.

[44] Y. Louzoun, The evolution of mathematical immunology, *Immunological Reviews*, **216** (2007), 9–20.

[45] T. Luzyanina, D. Roose, and G. Bocharov, Distributed parameter identification for a label-structured cell population dynamics model using CFSE histogram time-series data, *J. Math. Biol.* **59** (2009), 581–603.

[46] T. Luzyanina, M. Mrusek, J.T. Edwards, D. Roose, S. Ehl, and G. Bocharov, Computational analysis of CFSE proliferation assay, *J. Math. Biol.* **54** (2007), 57–89.

[47] T. Luzyanina, D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, and G. Bocharov, Numerical modelling of label-structured cell population growth using CFSE distribution data, *Theoretical Biology and Medical Modelling* **4** (2007), Published Online.

[48] A. B. Lyons, Analysing cell division in vivo and in vitro using flow cytometric measurement of CFSE dye dilution, *J. Immunological Methods*, **243** (2000), 147–154.

[49] A. B. Lyons, J. Hasbold and P.D. Hodgkin, Flow cytometric analysis of cell division history using diluation of carboxyfluorescein diacetate succinimidyl ester, a stably integrated fluorescent probe, *Methods in Cell Biology* **63** (2001), 375–398.

[50] A.B. Lyons and C.R. Parish, Determination of lymphocyte division by flow cytometry, *J. Immunol. Methods* **171** (1994), 131–137.

[51] G. Matera, M. Lupi and P. Ubezio, Heterogeneous cell response to topotecan in a CFSE-based proliferative test, *Cytometry A* **62** (2004), 118–128.

[52] J.A. Metz and O. Diekmann, The dynamics of physiologically structured populations, *Lecture Notes in Biomathematics* **68** (1986).

[53] R.E. Nordon, M. Nakamura, C. Ramirez, and R. Odell, Analysis of growth kinetics by division tracking, *Immunology and Cell Biology* **77** (1999), 523–529.

[54] C. Parish, Fluorescent dyes for lymphocyte migration and proliferation studies, *Immunology and Cell Biol.* **77** (1999), 499–508.

[55] B. Perthame, *Transport Equations in Biology,* Birkhauser Frontiers in Mathematics, Basel, 2007.

[56] B. Quah, H. Warren and C. Parish, Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester, *Nature Protocols* **2**:9 (2007), 2049–2056.

[57] P. Revy, M. Sospedra, B. Barbour, and A. Trautmann, Functional antigen-independent synapses formed between T cells and dendritic cells, *Nature Immunology* **2** (2001), 925–931.

[58] G.A. Sever and C.J. Wild, *Nonlinear Regression*, Wiley, Hoboken, 2003.

[59] T.E. Schlub, V. Venturi, K. Kedzierska, C. Wellard, P. Doherty, S.J. Turner, R.M. Ribeiro, P.D. Hodgkin, and M.P. Davenport, Division-linked differentiation can account for CD8+ T-cell phenotype in vivo, *Eur. J. Immunology* **39** (2009), 67–77.

[60] J. Sinko and W. Streifer, A New Model for Age-Size Structure of a Population, *Ecology* **48** (1967), 910–918.

[61] J.A. Smith and L. Martin, Do Cells Cycle?, *Proc. Natl. Acad. Sci* **70** (1973), 1263–1267.

[62] V.G. Subramanian, K.R. Duffy, M.L. Turner and P.D. Hodgkin, Determining the expected variability of immune responses using the cyton model, *J. Math. Biol.* **56** (2008), 861–892.

[63] B. Tummers, DataThief III. 2006 (http://datathief.org/)

[64] M.L. Turner, E.D. Hawkins, and P.D. Hodgkin, Quantitative regulation of B cell division destiny by signal strength, *J. Immunology* **181** (2009), 374–382.

[65] H. Veiga-Fernandez, U. Walter, C. Bourgeois, A. McLean, and B. Rocha, Response of naive and memory CD8+ T cells to antigen stimulation in vivo, *Nature Immunology* **1** (2000), 47–53.

[66] P.K. Wallace, J.D. Tario, Jr., J.L. Fisher, S.S. Wallace, M.S. Ernstoff, and K.A. Muirhead, Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution, *Cytometry A* **73** (2008), 1019–1034.

[67] C. Wellard, J. Markham, E.D. Hawkins, and P.D. Hodgkin, The effect of correlations on the population dynamics of lymphocytes, *J. Theoretical Biology* **264** (2010), 443–449.

[68] J.M. Witkowski, Advanced application of CFSE for cellular tracking, *Current Protocols in Cytometry* (2008), 9.25.1–9.25.8.

[69] A. Yates, C. Chan, J. Strid, S. Moon, R. Callard, A.J.T. George, and J. Stark, Reconstruction of cell population dynamics using CFSE, *BMC Bioinformatics* **8** (2007), Published Online.